Adaptive finite elements for optimally controlled elliptic variational inequalities of obstacle type

Dissertation zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät der
Universität Augsburg

vorgelegt von Alexandra Gaevskaya
Mai 2013

Erstgutachter:	Prof. Dr. Ronald H.W. Hoppe, Universität Augsburg
Zweitgutachter:	Prof. Dr. Malte A. Peter, Universität Augsburg
Drittgutachter:	Prof. Dr. Michael Hintermüller, Humboldt-Universität zu Berlin
Mündliche Prüfung:	24. Juli 2013

Contents

1 Introduction

This thesis is devoted to the problem of optimal control of a special kind of elliptic variational inequality, the so-called obstacle problem. This problem belongs to a wider class of problems called mathematical programs with complementarity constraints (MPCC), which, in turn, constitute a subclass of mathematical programs with equilibrium constraints (MPEC). MPECs have been subject to extensive studies in the recent past both in functional spaces [5, 7, 8, 9, 10, 33, 45, 41, 46, 48, 49] and in finite dimensions [54, 56, 24].

Roughly speaking, an MPEC is a constrained optimization problem where the set of constraints is described by a variational inequality. In the particular case of the obstacle problem, the variational inequality itself can be rephrased as another constrained optimization problem. The canonical example of an obstacle problem is the problem of deformation of a membrane under action of a force where the displacement of the membrane is restricted by an obstacle. The problem can be solved by minimizing an energy functional over all deformations compatible with the constraint. In terms of this canonical example, optimal control of the obstacle problem can be interpreted as a minimization problem where the force is varied in order to minimize the distance between the current state of the membrane and some desired shape. In this sense, the optimally controlled obstacle problem is a bilevel optimization problem.

In the analysis of optimally controlled elliptic obstacle problems one faces difficulties such as intrinsic non-convexity and non-differentiability that do not arise, for instance, for the problems of optimally controlled elliptic partial differential equations. Therefore the well-established approaches from optimal control and optimization theory are not applicable. Instead, one has to employ tools from non-smooth analysis, in particular concepts of generalized derivatives from which a variety of concepts for stationary points arise. Well known example of these are C(lark)- and S(trong)-stationary points. Furthermore, the non-convexity of the problem leads to the non-uniqueness of the optimal solutions.

In the recent past the adaptive finite element method has proved to be successfully applicable for the efficient numerical solution of a large number of classes of problems, such as elliptic variational inequalities [2, 11, 12, 13, 14, 17, 40, 42, 52, 53, 60] and unconstrained and constrained optimal control problems governed by partial differential equations [26, 25, 30, 32, 34, 35, 37, 38, 39, 47, 62]. Thus, it seems natural to investigate whether the adaptive finite element method could be equally successful employed for the solution of the optimally controlled obstacle problem. An adaptive approach based on goal oriented dual weighted residuals was suggested in [36]. In contrast, the main focus of this work lies in the development of an adaptive finite element scheme based on

a standard residual-type a posteriori error estimate. The following paragraph motivates the path taken by this thesis and summarizes the main results.

1.1 Abstract

In order to numerically solve the optimal control problem, a finite element discretization by means of continuous piecewise affine finite elements is introduced. In finite dimensions, a hierarchy of concepts of stationarity is available [56]. However, it is known that the optimally controlled obstacle problem always has S-stationary points (the strongest of the available characterizations of the minima) and a fortiori C-stationary points. Solutions of the optimal control problem discretized by finite elements can therefore be characterized by corresponding discrete C- or S-stationarity systems.

An analog of the hierarchy that occurs in the finite dimensional case can also be found for the optimally controlled elliptic variational inequalities in functional spaces [33, 45]. Due to the specifics of the functional space context, the classical stationary concepts obtain several interpretations. For instance, the above mentioned concept of C-stationarity in functional space is available in three different forms: as C-stationarity, almost C-stationarity, and ε-almost C-stationarity (cf. [33]). This is a consequence of the low regularity of Lagrangian multipliers in the associated stationarity systems. In order to make a well-grounded choice of a continuous stationarity system for later use in the a posteriori analysis, the question of convergence of discrete C-stationary points computed on sequences of uniformly refined meshes is considered. Under some additional assumptions, it is shown that the sequence of discrete C-stationary points converges to an almost C-stationary point of the continuous optimal control problem. This is the first main result of this thesis.

As the second main result, a residual-type a posteriori error estimator is suggested for which reliability and efficiency properties are established up to data oscillations and consistency errors due to mismatch in complementarity. The consistency errors are not “a posteriori”, that is, they depend on the exact solution of the problem and are therefore not computable. In order to deal with this problem, heuristically motivated fully computable estimates are suggested.

1.2 Outline

This work starts with an overview of properties of the obstacle problem in Chapter 2. A transition from the variational inequality formulation to a complementarity formulation with a slack variable (Lagrangian multiplier) is done. The slack variable is an element of the dual space and thus lacks pointwise interpretation. Local properties of the solutions of the obstacle problem with a special attention to the Lagrangian multiplier are then examined in detail. In particular, the notions of continuous active, inactive, strongly active, zero, and biactive sets are introduced and the basic properties of these sets are investigated.

In Chapter 3, the optimal control problem is introduced. The questions of the existence of solutions is addressed and a hierarchy of stationary points for the continuous setting is introduced. Finally, the local properties of the solutions of the continuous optimality systems with respect to the sets defined in the previous chapter are discussed.

A finite element discretization for the optimally controlled obstacle problem is introduced in Chapter 4. The question of local properties of the finite element approximations is addressed. At this point, a lot of attention is devoted to the definitions of the discrete sets in order to preserve the key properties of their analogs in the continuous setting. Further, C- and S-stationarity systems for the finite element discretization of the optimally controlled obstacle problem are derived with a subsequent comparative analysis of the local properties of the discrete and continuous stationary points. Finally, extensions of the discrete Lagrangian multipliers to the spaces where the Lagrangian multipliers from the continuous setting live are introduced.

Chapter 5 contains the proof of the convergence result for the discrete C-stationarity points. In the end of this chapter, the convergence result is discussed in view of the non-uniqueness of the solutions of the problem.

Chapter 6 is devoted to the derivation of a residual type a posteriori error estimate for the difference between the solutions of the discrete and the continuous optimal control problems. Reliability and efficiency of the residual-type error estimator up to consistency errors and data oscillations are verified and fully computable heuristic approximations for the consistency errors are introduced.

The adaptive finite element method based on the residual-type error estimate is introduced in Chapter 7. The individual steps of the algorithm are described in detail.

Chapter 8 offers a detailed documentation of the numerical results performed on a selection of test problems.

In Chapter 9, conclusions are drawn form the theoretical and numerical results of the thesis.

1.3 Notation and function spaces

In this work the standard notation is adopted for the set of real numbers $\mathbb{R}$, the set of natural numbers $\mathbb{N}$, and the set $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Bold letters will be used to denote vectors, matrices, and vector-valued functions. For $n, m \in \mathbb{N}$, let $\boldsymbol{M} \in \mathbb{R}^{n\times m}$ be a real-valued matrix with n rows and m columns and $\boldsymbol{v} \in \mathbb{R}^n$ a real-valued vector with n components. The elements of the matrix $\boldsymbol{M}$ are referred to as $(\boldsymbol{M})_{i.j}$, $i = 1, \dots, n$, $j = 1, \dots, m$ and the components of the vector $\boldsymbol{v}$ as $(\boldsymbol{v})_i$, $i = 1, \dots, n$. Furthermore, the zero vector in $\mathbb{R}^n$ is denoted by $\mathbf{0}_n$ and the zero element of $\mathbb{R}^{n\times m}$ by $\mathbf{0}_{nm}$. $\boldsymbol{I}_{nn} \in \mathbb{R}^{n\times n}$ stands for the identity matrix of the size $n \times n$. For a pair of vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ the notation $\boldsymbol{v} \cdot \boldsymbol{w} := \boldsymbol{v}^T \boldsymbol{w}$ is used for the scalar product of $\boldsymbol{v}$ and $\boldsymbol{w}$, while $\boldsymbol{v} * \boldsymbol{w}$ is used for the Hadamard (component-wise) product, i.e., $(\boldsymbol{v} * \boldsymbol{w})_i = (\boldsymbol{v})_i(\boldsymbol{w})_i$, $i = 1, \dots, n$. The inequality $\boldsymbol{w} \geq \boldsymbol{v}$ is understood in the component-wise sense, i.e., $(\boldsymbol{w})_i \geq (\boldsymbol{v})_i$ for all $i = 1, \dots, n$.

For any set $D \in \mathbb{R}^d$, $d \in \mathbb{N}$ the topological closure of D is denoted by $\overline{D}$ or $\mathrm{cl}(D)$, the interior of D by $\overset{\circ}{D}$ or $\mathrm{int}(D)$, and the boundary of D by ∂D. $\mathrm{meas}(D)$ and $|D|$ will denote, respectively, the d-dimensional Lebesgue measure of D and the area of D. Furthermore, Ω will denote an open bounded set in $\mathbb{R}^d$. Let vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^T \in \mathbb{N}_0^d$ with the length $|\boldsymbol{\alpha}| := \alpha_1 + \ldots + \alpha_d$ be referred to as *a multi-index*. For $\boldsymbol{x} = (x_1, \ldots, x_d)^T \in \Omega$, the component α_i of the multi-index α gives the number of differentiations of a function of $\boldsymbol{x}$ in x_i, $i \in 1, \ldots, d$. For an arbitrary subset $\omega \subset \Omega$, the characteristic function χ_ω of the set ω is defined by

$$\chi_\omega(x) := \begin{cases} 1, & x \in \omega \\ 0, & \text{else} \end{cases} .$$

The standard norm within a normed vector space X is denoted by $\|\cdot\|_X$, the dual space of X (the space of all linear continuous functionals on X) by X^*. For the associated duality pairing the notation $\langle \cdot, \cdot \rangle_{X^*,X}$ is employed. The nonnegative cone of X with respect to the ordering in X is denoted by X_+. The nonnegative cone of the dual space X_+^* is defined as

$$x^* \in X_+^* \quad \Leftrightarrow \quad \langle x^*, x \rangle \geq 0 \ \text{ for all } \ x \in X_+.$$

1.3.1 Spaces of continuous functions

For $m \in \mathbb{N}_0$, the spaces of continuous and continuously differentiable functions are defined as

$$C^m(\Omega) := \{v : \Omega \to \mathbb{R} \mid \partial^{\boldsymbol{\alpha}} v \text{ continuous for all } \boldsymbol{\alpha} \text{ with } |\boldsymbol{\alpha}| \leq m\},$$
$$C^m(\overline{\Omega}) := \{v : \overline{\Omega} \to \mathbb{R} \mid \partial^{\boldsymbol{\alpha}} v \text{ continuous for all } \boldsymbol{\alpha} \text{ with } |\boldsymbol{\alpha}| \leq m\},$$

where $\partial^{\boldsymbol{\alpha}} v$ is a partial derivative of v of order $\boldsymbol{\alpha}$

$$\partial^{\boldsymbol{\alpha}} v = \frac{\partial^{|\boldsymbol{\alpha}|} v}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.$$

Special attention receive the spaces of continuous functions $C(\Omega) := C^0(\Omega)$, $C(\overline{\Omega}) := C^0(\overline{\Omega})$, the space of functions whose partial derivatives of all orders are continuous

$$C^\infty(\Omega) := \bigcap_{m=0}^{\infty} C^m(\Omega),$$

and its subspace of functions with compact support

$$C_0^\infty(\Omega) := \{v \in C^\infty(\Omega) \mid \mathrm{supp}\, v \subset\subset \Omega\}.$$

Here $D \subset\subset \Omega$ signifies that the set D is compactly contained in Ω.
The class of Schwartz distributions $\mathcal{D}'(\Omega)$ consists of the linear continuous functionals

on $\mathcal{D}(\Omega) := C_0^\infty(\Omega)$. The reader is referred to [58] for more details.
By means of the representation

$$\langle v^*, v\rangle_{C(\overline{\Omega})^*, C(\overline{\Omega})} = \int_\Omega v(x) d\lambda(x), \quad v^* \in C(\overline{\Omega})^*,\, v \in C(\overline{\Omega})$$

the dual space $C(\overline{\Omega})^*$ of $C(\overline{\Omega})$ can be identified with the space of real regular Borel measures λ defined on $\overline{\Omega}$, which we denote by $\mathcal{M}(\overline{\Omega})$ (see, e.g., [3], [59], or [55]). Note that due to the compactness of $\overline{\Omega}$, $\mathcal{M}(\overline{\Omega})$ can be equivalently referred to as the space of Radon measures. In the sequel, the dual of $C(\overline{\Omega})$ will be addresses as $\mathcal{M}(\overline{\Omega})$, the notation $\langle\langle \cdot, \cdot \rangle\rangle$ will be used for the corresponding duality pairing.

1.3.2 Lebesgue spaces

For $1 \le p \le \infty$, the standard Lebesgue spaces are defined as follows

$$L^p(\Omega) := \{v : \Omega \to \mathbb{R} \mid v \text{ is Lebesgue measurable, } \|v\|_{L^p(\Omega)} < \infty\},$$

where

$$\|v\|_{L^p(\Omega)} = \left(\int_\Omega |v(x)|^p \, dx \right)^{1/p}, \; 1 \le p < \infty, \quad \text{and } \|v\|_{L^\infty(\Omega)} = \operatorname*{ess\,sup}_{x \in \Omega} |v(x)|.$$

For every $1 \le p \le \infty$, the space $L^p(\Omega)$ endowed with the norm $\|\cdot\|_{L^p(\Omega)}$ is a Banach space. For the case $p = 2$ we may define a scalar product

$$(v, w)_{L^2(\Omega)} := \int_\Omega v(x) w(x) \, dx,$$

giving the space $L^2(\Omega)$ the structure of a Hilbert space. Further, we can introduce the set of locally integrable functions

$$L^1_{loc}(\Omega) := \{v : \Omega \to \mathbb{R} \mid v \in L^1(\Omega') \text{ for every open } \Omega' \text{ with } \Omega' \subset\subset \Omega\}.$$

1.3.3 Sobolev spaces

A function $v \in L^1_{loc}(\Omega)$ has a *weak* (or distributional) *partial derivative* of order α if there exists $D^{\boldsymbol{\alpha}} v := w \in L^1_{loc}(\Omega)$ satisfying

$$\int_\Omega w(x)\, \varphi(x)\, dx = (-1)^{|\boldsymbol{\alpha}|} \int_\Omega v(x)\, \partial^{\boldsymbol{\alpha}} \varphi(x)\, dx, \quad \forall\, \varphi \in C_0^\infty(\Omega). \tag{1.3.1}$$

For any $m \in \mathbb{N}_0$ and $1 \le p \le \infty$, we define the Sobolev spaces

$$W^{m,p}(\Omega) := \{v \in L^p(\Omega) \mid D^{\boldsymbol{\alpha}} v \in L^p(\Omega) \text{ for } 0 \le |\boldsymbol{\alpha}| \le m\}$$

and

$$W_0^{m,p}(\Omega) := \{v \in W^{m,p}(\Omega) \mid \exists v_n \in C_0^\infty(\Omega) \text{ with } \lim_{n\to\infty} \|v - v_n\|_{W^{m,p}(\Omega)} = 0\}$$

endowed with the norm

$$\|v\|_{W^{m,p}(\Omega)} = \left(\sum_{0\le|\alpha|\le m} \|D^\alpha v\|^p_{L^p(\Omega)} \right)^{1/p}, \ 1 \le p < \infty,$$
$$\|v\|_{W^{m,\infty}(\Omega)} = \max_{0\le|\alpha|\le m} \|D^\alpha v\|_{L^\infty(\Omega)},$$

as well as the seminorm

$$|v|_{W^{m,p}(\Omega)} = \left(\sum_{|\alpha|=m} \|D^\alpha v\|^p_{L^p(\Omega)} \right)^{1/p}, \ 1 \le p < \infty,$$
$$|v|_{W^{m,\infty}(\Omega)} = \max_{|\alpha|=m} \|D^\alpha v\|_{L^\infty(\Omega)}.$$

The Sobolev spaces $W^{m,p}(\Omega)$ and $W_0^{m,p}(\Omega)$ with corresponding norms are Banach spaces.
In the case $p = 2$, it is customary to use the notation $H^m(\Omega) := W^{m,2}(\Omega)$, $H_0^m(\Omega) := W_0^{m,2}(\Omega)$, for the spaces and $\|\cdot\|_{m,\Omega} := \|\cdot\|_{W^{m,2}(\Omega)}$, $|\cdot|_{m,\Omega} := |\cdot|_{W^{m,2}(\Omega)}$ for the norms and the seminorms. $H^m(\Omega)$ and $H_0^m(\Omega)$ are Hilbert spaces with the scalar product

$$(v, w)_{m,\Omega} := \sum_{|\alpha|\le m} (D^\alpha v, D^\alpha w)_{0,\Omega}.$$

The identity $H^0(\Omega) = L^2(\Omega)$ motivates the notation $(\cdot,\cdot)_{0,\Omega} := (\cdot,\cdot)_{L^2(\Omega)}$ and $\|\cdot\|_{0,\Omega} := \|\cdot\|_{L^2(\Omega)}$ for the scalar product and the associated norm in the space $L^2(\Omega)$. In the sequel, for any subset D of Ω we refer to $L^2(D)$ as the Hilbert space with the inner product $(\cdot,\cdot)_{0,D}$ and the associated norm $\|\cdot\|_{0,D}$. We further refer to $H^m(D)$, $m \in \mathbb{N}$ as the Sobolev spaces with the norm $\|\cdot\|_{m,D}$ and seminorm $|\cdot|_{m,D}$.

By $H^{-m}(\Omega)$, $m \in \mathbb{N}$ we denote the dual space of $H_0^m(\Omega)$. In what follows, let us reserve the notation $\langle\cdot,\cdot\rangle$ for the duality pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$.

For a more detailed study of Sobolev spaces, the reader is referred to the monographs [3], [1], [22], or [29].

1.3.4 Traces and extensions

As $\partial\Omega$ has a d-dimensional Lebesgue measure zero, to prescribe boundary values to Sobolev functions one needs the concepts of Lebesgue spaces on $\partial\Omega$ (we refer, e.g., to [3]) and of the trace operator. The latter is defined for a special class of Lipschitz domains (cf. [29] or [11]).

Definition 1.3.1. *(Lipschitz domain, domain of the class $C^{1,1}$)*

- *A function $f : \mathbb{R}^n \supset D \to \mathbb{R}^m$ is called Lipschitz continuous if there exists $L > 0$ such that $\|f(x) - f(y)\|_{\mathbb{R}^m} \leq L\|x - y\|_{\mathbb{R}^n}$ for all $x, y \in D$.*
- *A hypersurface in $\mathbb{R}^d$ is a graph if it can be represented in the form $x_k = f(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_d)$, with $1 \leq k \leq d$ and some suitable domain in $\mathbb{R}^{d-1}$.*
- *A domain Ω is called Lipschitz provided that it lies on one side of the boundary $\partial\Omega$ and that for every $x \in \partial\Omega$ there exists a neighborhood of x in $\partial\Omega$ that can be represented as a graph of a Lipschitz continuous function. In this case it is also common to say that Ω is of the class $C^{0,1}$, where $C^{0,1}$ is the standard notation for Lipschitz continuous functions considered as special instance of a more general class of the so-called Hölder continuous functions (see, e.g., [1]).*
- *We say that Ω is of the class $C^{1,1}$ if for every $x \in \partial\Omega$ there exists a neighborhood of x in $\partial\Omega$ that can be represented as the graph of a one time continuously differentiable function whose first derivative is a Lipschitz continuous function.*

The following trace theorem is stated, e.g., in [3].

Theorem 1.3.2. *(Trace theorem)*
Let Ω be Lipschitz and let $1 \leq p \leq \infty$. There exists a unique linear continuous mapping

$$\tau \ : \ W_p^1(\Omega) \to L^p(\partial\Omega)$$

with

$$\tau v = v|_{\partial\Omega}, \ \forall\, v \in W_p^1(\Omega) \cap C(\overline{\Omega}).$$

In the special case of $\partial\Omega$ Lipschitz, $H_0^1(\Omega)$ can be identified with those elements of $H^1(\Omega)$ whose traces on $\partial\Omega$ are zero.

Theorem 1.3.3. *(Extension theorem (cf. [23], p.135))*
Suppose Ω is Lipschitz. For any bounded open set $\widetilde{\Omega}$ such that $\Omega \subset\subset \widetilde{\Omega}$ there exists a bounded linear operator

$$E : H^1(\Omega) \to H^1(\mathbb{R}^d) \tag{1.3.2}$$

such that for each $v \in H^1(\Omega)$:

(i) $Ev = v$ *a.e. in* Ω,

(ii) $supp(Ev) \subseteq \widetilde{\Omega}$,

(iii) $\|Ev\|_{1,\mathbb{R}^d} \leq C\|v\|_{1,\Omega}$ *with the constant C depending only on Ω and $\widetilde{\Omega}$.*

2 Obstacle problem

The first section of this chapter provides an overview of the relevant parts of the theory of general elliptic variational inequalities of the first kind in Hilbert space. In Section 2.2, these results are narrowed to the specific case of obstacle problems with solutions in $H_0^1(\Omega)$. The rest of the chapter is devoted to the properties of the solutions of the obstacle problem, which require a somewhat deeper insight into the properties of the elements of the dual space $(H_0^1(\Omega))^*$.

2.1 Elliptic variational inequalities of the first kind in Hilbert spaces

Let V be a real Hilbert space and V^* its dual space. By $\|\cdot\|_V$ we can denote the norm in V and by $\langle\cdot,\cdot\rangle$ the duality pairing between V and V^*. Further, let us introduce a *bilinear* form

$$a(\,\cdot\,,\,\cdot\,) : V \times V \to \mathbb{R},$$

that is assumed to be *bounded*, i.e.,

$$\exists\ \Gamma_a > 0 \quad : \quad |a(v,w)| \ \leq\ \Gamma_a \|v\|_V \|w\|_V, \quad \forall v, w \in V, \tag{2.1.1a}$$

and *coercive*, i.e.,

$$\exists\ \gamma_a > 0 \quad : \quad a(v,v) \ \geq\ \gamma_a \|v\|_V^2, \quad \forall v \in V. \tag{2.1.1b}$$

A bilinear form coercive with respect to V is also called V*-elliptic* (see, e.g., [27]).

As it is pointed out, e.g., in [44], for every bounded bilinear form there exists a bounded linear operator $A \ : V \to V^*$ such that

$$a(v,w) = \langle Av, w\rangle, \quad \forall v, w \in V. \tag{2.1.2}$$

And vice versa, every bounded linear operator defines a bounded bilinear form via the relation (2.1.2). Clearly, the properties of the bilinear form (2.1.1a), (2.1.1b) are passed on to the operator A as

$$\begin{aligned} \|Av\|_{V^*} &\leq \Gamma_a \|v\|_V, \quad \forall v \in V, \\ \langle Av, v\rangle &\geq \gamma_a \|v\|_V^2, \quad \forall v \in V. \end{aligned}$$

We refer to $\|A\|_{\mathcal{L}(V,V^*)}$ as the operator norm of A, i.e.,

$$\|A\|_{\mathcal{L}(V,V^*)} = \sup_{v\in V\setminus\{0\}} \frac{\|Av\|_{V^*}}{\|v\|_V}.$$

and observe that $\|A\|_{\mathcal{L}(V,V^*)} \leq \Gamma_a$.

A coercive bilinear form $a(\cdot,\cdot)$ defines a norm $a(\cdot,\cdot)^{1/2}$ on V equivalent to $\|\cdot\|_V$. If in addition the bilinear form $a(\cdot,\cdot)$ is *symmetric*, i.e.,

$$a(v,w) = a(w,v), \quad \forall v,w \in V,$$

it defines a scalar product on V thus making $(V, a(\cdot,\cdot))$ a Hilbert space.

Let K be a convex and closed subset of V and $g \in V^*$. Then *the elliptic variational inequality of the first kind* is a problem of the form:

find $y \in K$ such that

$$a(y, y-v) \;\leq\; \langle g, y-v\rangle\,, \quad \forall v \in K. \tag{2.1.3}$$

Definition 2.1.1. *(Control-to-state mapping)*
Let us introduce the solution operator corresponding to the variational inequality (2.1.3) *as*

$$\begin{aligned} \mathcal{S}: \quad V^* &\longrightarrow V, \\ g &\longmapsto y. \end{aligned} \tag{2.1.4}$$

In the context of the optimal control of problem (2.1.3) *this operator will be referred to as the control-to-state mapping.*

The following theorem states the existence and the uniqueness of the solution of (2.1.3).

Theorem 2.1.2. *Let $a(\,\cdot\,,\,\cdot\,)$ be a bounded and coercive bilinear form on V, and $K \subset V$ be a closed convex set. For each $g \in V^*$ the variational inequality* (2.1.3) *admits a unique solution $y \in V$. Additionally, the mapping $\mathcal{S}$ is Lipschitz continuous, i.e., if y_1 and y_2 are two solutions of problem* (2.1.3) *corresponding to right hand sides $g_1, g_2 \in V^*$, then*

$$\|y_1 - y_2\|_{1,\Omega} \leq \frac{1}{\gamma_a}\|g_1 - g_2\|_{V^*},$$

where $\gamma_a > 0$ is the coercivity constant defined in (2.1.1b).

Proof. The proof can be found, e.g., in [44], p. 26. □

If, in addition to the assumptions of Theorem 2.1.2, the bilinear form $a(\,\cdot\,,\,\cdot\,)$ is symmetric, the variational inequality (2.1.3) is equivalent to the minimization problem

$$\min_{y\in K} I(y), \quad I(y) := \frac{1}{2}a(y,y) - \langle g, y\rangle, \tag{2.1.5}$$

see, e.g., [28].

2.2 Obstacle problem in $H_0^1(\Omega)$

One of the classical examples of a variational inequality of the first kind is given by the so-called *obstacle problem*. This problem will determine the set of constraints in the optimal control problem we will consider later. Let us start with a formal definition of the obstacle problem.

Let $\Omega \subset \mathbb{R}^2$ be a domain with boundary $\Gamma := \partial\Omega$. Assume that

$$\Omega \text{ – open, bounded, polygonal Lipschitz domain.} \tag{2.2.1}$$

The thus defined Ω may be non-convex, possibly with reentrant corners. Whereas (2.2.1) will be our basic assumption on Ω, some results will require a stronger assumption

$$\Omega \text{ – open, bounded, and either convex and polygonal or of the class } C^{1,1}. \tag{2.2.2}$$

In the literature devoted to the optimal control of variational inequalities the authors often rely on the latter stronger assumption. It should be emphasized that in this work it plays a secondary role.

From here on, we set $V := H_0^1(\Omega)$. Abusing the general notation for normed spaces from Section 1.3, we set

$$V_+ := \{v \in V \,|\, v(x) \geq 0 \text{ for a.a. } x \in \Omega\},$$

i.e., V_+ will denote the nonnegative cone of V with respect to the ordering of $L^2(\Omega)$ We refer to V_+^* as the nonnegative cone of V^* associated with V_+, more precisely,

$$V_+^* := \{\lambda \in V^* \,|\, \langle \lambda, v\rangle \geq 0,\ \forall\ v \in V_+\}.$$

Finally, in agreement with the notation of Section 1.3, we define $C_+(\overline{\Omega})$ and $\mathcal{M}_+(\overline{\Omega})$ as the nonnegative cones of $C(\overline{\Omega})$ and its dual $\mathcal{M}(\overline{\Omega})$,

$$\begin{aligned} C_+(\overline{\Omega}) &:= \{v \in C_+(\overline{\Omega}) \,|\, v(x) \geq 0,\ x \in \overline{\Omega}\},\\ \mathcal{M}_+(\overline{\Omega}) &:= \{\lambda \in \mathcal{M}_+(\overline{\Omega}) \,|\, \langle\langle \lambda, v\rangle\rangle \geq 0,\ \forall\ v \in C_+(\overline{\Omega})\}. \end{aligned}$$

Although the subsequent analysis can be carried out for a general second order elliptic differential operator in divergence form, in the sequel we will restrict ourselves to the special case where the operator A introduced in the previous section equals $-\Delta$, i.e., we consider the bounded, coercive, symmetric bilinear form

$$a(v, w) := (\nabla v, \nabla w)_{0,\Omega},\ v, w \in V. \tag{2.2.3}$$

Additionally, for $D \subseteq \Omega$ we will use the notation $a_D(v, w) := (\nabla v, \nabla w)_{0,D}$, $v, w \in V$. For a given function

$$\psi \in V \tag{2.2.4}$$

we set

$$K := \{v \in V \,|\, v \leq \psi \text{ a.e. in } \Omega\}. \tag{2.2.5}$$

The thus defined set K is closed, convex, and non-empty. Further, let $g \in L^2(\Omega)$ be given.

With the notation introduced above, we can formulate the *obstacle problem*:

find $y \in K$ such that

$$\min_{y \in K} I(y), \quad I(y) = \frac{1}{2} a(y,y) - (g,y)_{0,\Omega}. \tag{2.2.6}$$

The model problem (2.2.6) embodies a number of various physical processes, one of which is the problem of deformation of a membrane subject to a unilateral constraint. The following example briefly explains the meaning of separate elements of (2.2.6) in application to this problem. Note that more details on this and other applications can be found, e.g., in [55] or [21].

Example 2.2.1. *(Deformation of a membrane constrained by an obstacle)*
The solution y of (2.2.6) can be interpreted as the vertical displacement of an elastic membrane occupying the region Ω and being fixed at the boundary Γ. The membrane lies under the obstacle ψ and is being pushed towards the obstacle by a vertical force $G = gt$ (t - the uniform tension of the unperturbed membrane). The functional $I(y)$ describes the total potential energy of the membrane for a given displacement $y \in K$. The total potential energy equals the potential energy of the deformation given by the increase in the area of the surface of the membrane $\frac{1}{2}a(y,y)$, minus the work done by the external forces during the actual displacement $(g,y)_{0,\Omega}$. According to the principle of minimum total potential energy, among all elements of K (functions of finite energy of deformation satisfying the homogeneous boundary conditions and lying under the obstacle ψ) the displacement function y^* corresponding to the position of equilibrium of the system will deliver the minimum to the functional I. This is the function we are looking for in (2.2.6).

Remark 2.2.2. One could consider a more general obstacle function

$$\psi \in H^1(\Omega), \text{ with } \tau\psi \geq 0 \text{ a.e. on } \Gamma,$$

where τ is the trace operator. This, however, might cause some changes in the complementarity formulation that will be introduced in the next paragraph.

2.2.1 Alternative formulations

Due to the symmetry of the bilinear form, the minimization problem (2.2.6) is equivalent to the variational inequality of the first kind

find $y \in K$ such that

$$a(y, y - v) \;\leq\; (g, y - v)_{0,\Omega}\,, \quad \forall v \in K, \tag{2.2.7}$$

which by Theorem 2.1.2 admits a unique solution.

It is convenient to introduce a slack variable $\sigma \in V^*$ (the residual of the inequality (2.2.7)),

$$\langle \sigma, v \rangle := (g, v)_{0,\Omega} - a(y, v).$$

The functional σ can also be viewed as a Lagrangian multiplier associated with the constraint $y \in K$ in the minimization problem (2.1.5). With σ, one can reformulate problem (2.2.7) as

$$a(y, v) = (g, v)_{0,\Omega} - \langle \sigma, v \rangle, \quad v \in V, \tag{2.2.8a}$$
$$y \in K, \quad \langle \sigma, y - v \rangle \geq 0, \quad v \in K. \tag{2.2.8b}$$

Condition (2.2.8b) is the inequality (2.2.7) written in terms of the residual σ. Testing (2.2.8b) subsequently with $v_1 = \psi \in K$ and $v_1 = 2y - \psi \in K$, we obtain $\langle \sigma, y - \psi \rangle = 0$. Consequently, $\langle \sigma, y - v \rangle = \langle \sigma, \psi - v \rangle$, and, therefore, (2.2.8b) implies $\langle \sigma, v \rangle \geq 0$ for $v \in V_+$. Thus the equivalence

$$(2.2.8b) \iff y \in K,\;\; \sigma \in V_+^*\,,\;\; \langle \sigma, y - \psi \rangle = 0.$$

Herewith, we can equivalently reformulate the obstacle problem as

find $(y, \sigma) \in V \times V^*$ such that

$$a(y, v) = (g, v)_{0,\Omega} - \langle \sigma, v \rangle, \quad v \in V, \tag{2.2.9a}$$
$$y \in K,\;\; \sigma \in V_+^*\,,\;\; \langle \sigma, y - \psi \rangle = 0. \tag{2.2.9b}$$

Condition (2.2.9b) is often referred to as the linear complementarity constraint and problem (2.2.9) as the problem with complementarity constraints. This name is due to the specific structure of this constraint: the components $\psi - y$ and σ are complementary in the sense that they are both nonnegative and, as their pairing must be zero, they cannot be strictly positive simultaneously.

2.2.2 Structure of the feasible set

The feasible set of problem (2.2.9) is given by

$$\mathcal{D} := \{(y, \sigma) \in V \times V^* \mid y \in K,\ \sigma \in V_+^*,\ \langle \sigma, y - \psi \rangle = 0\,\}.$$

This set is nonempty as the obstacle problem has a unique solution.

Let us have a closer look at set $\mathcal{D}$. Firstly, we observe that the constraint $\langle \sigma, y - \psi \rangle = 0$ is bilinear, and therefore non-convex. Thus, the feasible set $\mathcal{D}$ is non-convex. Moreover, as illustrated in Figure 2.1, the interior of this set is empty. Secondly, the point on the graph $(y - \psi, \sigma) = (0, 0)$ suggests the presence of the so-called "biactive set" $\mathcal{B}$ which can roughly be characterized as $\mathcal{B} := \{"y - \psi = 0"\} \cap \{"\sigma = 0"\}$. The meaning of this definition in view of the low regularity of σ will be discussed in Section 2.3.1. The biactive set is known to yield auxiliary challenges in the analysis and numerical solution of the optimal control problem governed by (2.2.9).

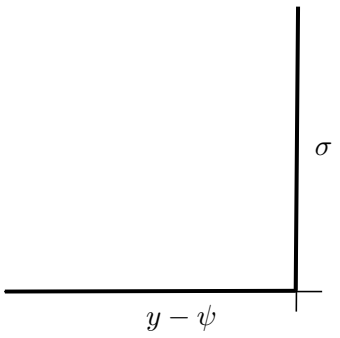


Figure 2.1: The feasible set $\mathcal{D}$

2.2.3 Control-to-state mapping

The control-to-state mapping $\mathcal{S}$ (cf. Definition 2.1.1) can now be written as

$$\begin{aligned} \mathcal{S}: \quad V^* &\longrightarrow V \times V^*, \\ g &\longmapsto (y, \sigma). \end{aligned} \tag{2.2.10}$$

The properties of $\mathcal{S}$ in view of the observations of the previous paragraph, including the differentiability results proved in [48] and revised in [46] are summarized in the following proposition.

Proposition 2.2.3.

(*i*) *$\mathcal{S}$ is a non-linear (bilinear), non-convex, Lipschitz continuous mapping.*

(*ii*) *$\mathcal{S}$ is conically differentiable. If at some g there holds $\mathcal{B} = \varnothing$, then $\mathcal{S}$ is Gâteaux differentiable at g.*

Remark 2.2.4. In [48], the term "conical derivative" is reserved for positively homogeneous directional derivatives of continuous mappings. For the definition of the notion of convexity for mappings in Banach spaces see, e.g., [59], p. 245.

2.2.4 Regularity of the solution

Lemma 2.2.5. *Let $a(\cdot,\cdot)$ be a bounded and coercive bilinear form on V, and K be defined by* (2.2.6)*. For $g \in L^2(\Omega)$ and $\psi \in L^\infty(\Omega) \cap V$, the unique solution y belongs to $L^\infty(\Omega) \cap V$, and the mapping $\mathcal{S}$ is Lipschitz continuous from $L^2(\Omega)$ to $L^\infty(\Omega)$.*

Proof. See [46], p. 7. □

Higher regularity of the solution can be obtained under stronger assumptions on the domain Ω and the obstacle function ψ.

Lemma 2.2.6. *Additionally to the requirements of Lemma 2.2.5, assume that* (2.2.2) *holds true and the obstacle ψ fulfills*

$$\psi \in V \cap H^2(\Omega). \tag{2.2.11}$$

For $g \in L^2(\Omega)$ the unique solution (y,σ) of (2.2.9) *belongs to $V \cap H^2(\Omega) \times L^2(\Omega)$.*

Proof. See [46], p. 7. □

Remark 2.2.7. If the assumptions of Lemma 2.2.6 are satisfied, the complementarity condition (2.2.9b) can be rephrased as

$$\psi - y \in V_+, \;\; \sigma \in (L^2(\Omega))_+ \,, \;\; (\sigma, y-\psi)_{0,\Omega} = 0. \tag{2.2.12}$$

Moreover, according to the Sobolev embedding theorem (see, e.g., [3], p. 333), the state function is also a continuous function, i.e., $y \in C(\overline{\Omega})$.

2.3 Properties of V^*-functionals

In this section we take a closed look at selected properties of the space V^*.

Based on the Riesz representation theorem, we can identify the space $L^2(\Omega)$ with its dual $(L^2(\Omega))^*$. It is convenient as the identification is based on the L^2-scalar product. Technically, we could do the same for V. Indeed, V is a Hilbert space, thus, by the Riesz theorem, for any $\lambda \in V^*$ there exists a unique $g_\lambda \in V$ satisfying

$$(g_\lambda, v)_{1,\Omega} = \langle \lambda, v \rangle, \quad \forall\, v \in V. \tag{2.3.1}$$

Due to analytical inconvenience (the H^1-scalar product in the determining relation (2.3.1)), this practice is conventionally not applied to V. Instead, the so-called *Gelfand triple* is used to connect V with its dual V^* (see, e.g., [63]). It is based on the embeddings

$$i_1 \;:\; V \to L^2(\Omega), \tag{2.3.2}$$

$$i_2 \;:\; L^2(\Omega) \to V^*, \tag{2.3.3}$$

where the first operator can be trivially defined via the relation

$$\forall\ v \in V, \quad i_1(v) := v,$$

whereas the second one maps every element $w \in L^2(\Omega)$ to $i_2(w) \in V^*$ by means of the relation

$$\langle i_2(w), v\rangle := (w, i_1(v))_{0,\Omega} = (w, v)_{0,\Omega}, \quad \forall\ v \in V,$$

and relies on the fact that the functional $v \mapsto (w, v)_{0,\Omega} \in \mathbb{R}$ for $v \in V$ is a linear continuous functional on V, i.e., every element $w \in L^2(\Omega)$ can be understood as an element of V^*. This yields a Gelfand triple

$$V \hookrightarrow_{i_1} L^2(\Omega) \hookrightarrow_{i_2} V^*,$$

with each space (continuously) embedded and dense in the subsequent space (see [63], p. 253).

Remark 2.3.1. The space $L^2(\Omega)$ is richer then V. A functional $\lambda \in V^*$ can be continuously extended to $(L^2(\Omega))^*$ only if there exists such $v_\lambda \in L^2(\Omega)$ that the action of λ can be expressed in the form $\langle \lambda, v\rangle = (v_\lambda, v)_{0,\Omega}$, $\forall\ v \in V$. In general, by the Riesz representation theorem, for any $\lambda \in V^*$ there exist functions $\lambda_\alpha \in L^2(\Omega)$, $0 \le |\alpha| \le 1$, such that

$$\langle \lambda, v\rangle = \sum_{0\le|\alpha|\le 1} (\lambda_\alpha, D^\alpha v)_{0,\Omega}, \quad \forall\, v \in V. \tag{2.3.4}$$

The reader is referred, e.g., to [22].

Remark 2.3.2. Any $\lambda \in V^*$ can be evaluated on functions from $\mathcal{D}(\Omega)$. Thus, elements of V^* are Schwartz distributions.

We can introduce a norm on V^* according to

$$\|\lambda\|_{-1,\Omega} := \sup_{v\in V\setminus\{0\}} \frac{|\langle \lambda, v\rangle|}{\|v\|_{1,\Omega}}.$$

2.3.1 Local properties

The main concern of this paragraph is the characterization of local properties of V^*-functionals (properties that hold on subsets of Ω). One point of interest is, for example, the definition of support for the elements of V^*. Let us first recall the definitions of the support of the continuous and the $L^2(\Omega)$ functions, as well as the support of the distributions (the latter can be found, e.g., in [58] or [63]).

Definition 2.3.3. *(Support of functions)*

(*i*) *For* $v \in C(\Omega)$ $(v \in \mathcal{D}(\Omega))$, $\operatorname{supp}(v) := \overline{\{x \in \Omega \,|\, v(x) \neq 0\}} = \Omega \setminus \mathcal{O}_v$, $\mathcal{O}_v :=$ *{ the maximal open set where* v *vanishes };*

(*ii*) *for* $v \in L^2(\Omega)$ $(v \in V)$, $\mathcal{O}_v := \{$ *the maximal open set where* $v = 0$ *holds a.e.* $\}$, $\operatorname{supp}(v) := \Omega \setminus \mathcal{O}_v$.

As the functions of V are subject to the L^2-ordering, the definition for $L^2(\Omega)$ is also suited for V.

Unlike the continuous or the $L^2(\Omega)$ functions, distributions do not admit a pointwise interpretation. Thus, before defining the notion of support, we must specify what it means for a distribution to be zero on a subset of Ω. A natural way to get information about local properties of a functional is to evaluate it on suitably chosen subspaces of the argument space. The following definition was adopted from [58].

Definition 2.3.4. *For* $T \in \mathcal{D}'(\Omega)$ *and* $\omega \subseteq \Omega$ *open, we say that* $T = 0$ *on* ω *if its effect on test functions with compact support in* ω *is zero, i.e.,*

$$T(v) = 0, \quad \forall\, v \in \mathcal{D}(\Omega) \text{ such that } \operatorname{supp}(v) \subseteq \omega.$$

With this notion at hand, the definition of the support of a distribution is straightforward. Being based on the distribution-specific notion of "$T = 0$ on ω", it otherwise resembles Definition 2.3.3.

Definition 2.3.5. *(Support of distributions) For* $T \in \mathcal{D}'(\Omega)$, $\mathcal{O}_T := \{$ *the maximal open set where* T *vanishes* $\}$, $\operatorname{supp}(T) := \Omega \setminus \mathcal{O}_T$.

We observe that the elements of both, V^* and $\mathcal{D}'(\Omega)$, do not admit a pointwise interpretation. Moreover, as pointed out in Remark 2.3.2, the elements of V^* are distributions. Thus, it seems natural to consider the localization of distributions as a prototype for the definition of analogous properties for V^*-functionals. Though, as it will be outlined in Remark 2.3.8, the question of a proper choice of test subspaces for the localization of the elements of V^* appears to be rather delicate.

For $\omega \subseteq \Omega$, we introduce a subspace of localized test functions $V_{\omega,0} \subseteq V$ according to

$$V_{\omega,0} := \{v \in V \,|\, v = 0 \text{ a.e. in } \Omega \setminus \omega,\ v|_\omega \in H_0^1(\omega)\}.$$

In the following definitions, it is specified what it means for a functional from V^* to be zero or to have a sign on an open subset of Ω and define the notion of support for elements of V^*.

Definition 2.3.6. *For* $\lambda \in V^*$ *and* $\omega \subseteq \Omega$ *open, we say that*

(*i*) $\lambda = 0$ *on* ω *if* $\langle \lambda, v\rangle = 0$, $\forall\, v \in V_{\omega,0}$;

(*ii*) $\lambda \geq 0$ (≤ 0) *on* ω *if* $\langle \lambda, v\rangle \geq 0$ (≤ 0), $\forall\, v \in V_{\omega,0} \cap V_+$;

Note that a similar characterization is used in [60].

Definition 2.3.7. *(Zero set and support of* V^**-functionals) For* $\lambda \in V^*$, *the zero set of* λ *is given by* $\mathcal{O}_\lambda := \{$ *the maximal open set where* λ *vanishes* $\}$, *whereas* $\operatorname{supp}(\lambda) := \Omega \setminus \mathcal{O}_\lambda$.

Finally, for $\omega \subseteq \Omega$ we can define a local norm (cf. [11])

$$\|\lambda\|_{-1,\omega} := \sup_{v \in V_{\omega,0}\setminus\{0\}} \frac{|\langle \lambda, v\rangle|}{\|v\|_{1,\Omega}} = \sup_{v \in V_{\omega,0}\setminus\{0\}} \frac{|\langle \lambda, v\rangle|}{\|v\|_{1,\omega}},$$

that evidently satisfies

$$\langle \lambda, v\rangle \leq \|\lambda\|_{-1,\omega}\|v\|_{1,\omega}, \quad \forall\, v \in V_{\omega,0}. \tag{2.3.5}$$

Remark 2.3.8. We recall the classification of concepts for the localization of V^*-functionals introduced in [33]. For that we need to introduce another type of localized test functions,

$$V_\omega := \{v \in V \,|\, v = 0 \text{ a.e. in } \Omega \setminus \omega\} = \{v \in V \mid \operatorname{supp}(v) \subseteq \overline{\omega}\}.$$

Obviously, $V_{\omega,0} \subset V_\omega$ for any $\omega \subset \Omega$. If ω is Lipschitz, $V_{\omega,0} = V_\omega$ (see, e.g., [45], Lemma A.1). As Ω is a Lipschitz domain, $V_{\Omega,0} = V_\Omega = V$. In [33], several realizations of the statement

$$\lambda = 0 \text{ on } \omega,\ \omega \subseteq \Omega \text{ - open}, \tag{2.3.6}$$

for $\lambda \in V^*$ are suggested, namely,

(a) $\langle \lambda, v\rangle = 0,\ \forall\, v \in V_\omega$,

(b) $\langle \lambda, v\rangle = 0,\ \forall\, v \in V_{\omega,0}$,

(c) $\forall \varepsilon > 0,\ \exists\, U_\varepsilon \subseteq \omega$ with $\operatorname{meas}(\omega \setminus U_\varepsilon) \leq \varepsilon$ such that $\langle \lambda, v\rangle = 0,\ \forall\, v \in V_{U_\varepsilon}$.

Evidently, the latter concepts admit the hierarchy $(a) \Rightarrow (b) \Rightarrow (c)$. Moreover, $(a) = (b)$ if ω is Lipschitz. It should be emphasized that definition (c) does not impose any restrictions on the shape and the location of U_ε within ω. U_ε is free to occupy any subset of ω with the only restriction $\operatorname{meas}(\omega \setminus U_\varepsilon) <= \varepsilon$. Thus, e.g., $U_\varepsilon = \omega$ would also be an admissible choice. In Figure 2.2, an example of a configuration $\omega \subset \Omega$ and three different shapes of supports of test functions admissible in definitions (c) are shown.

The uncertainty in the interpretation of (2.3.6) arises due to the low regularity of the V^* functionals. Was there a pointwise interpretation of λ, properties $(a) - (c)$ would coincide. Definition 2.3.6 (i) obviously coincides with (b). From now on, we will work with concept (b) as:

- it aligns with the standard concept of localization of distributions;
- it coincides with the natural definition of "$\lambda = 0$ in Ω" ($\langle \lambda, v\rangle = 0\ \forall\, v \in V$) if Ω is non-Lipschitz.

However, it is not asserted that concepts (a) and (c) should be deprived of attention.

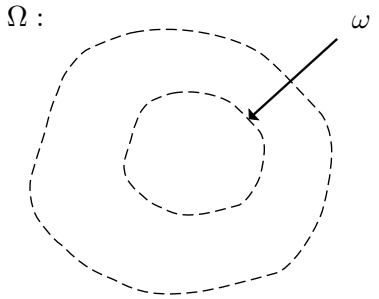
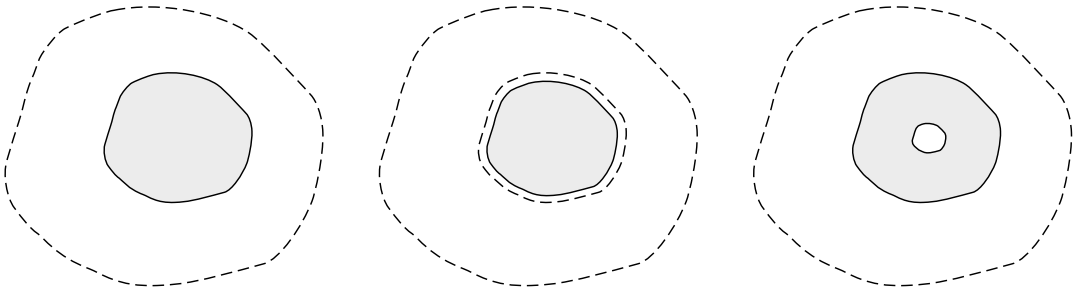


Figure 2.2: An example of a configuration $\omega \subset \Omega$ with an illustration of three possible shapes of supports of the test functions from definition (c). Supports are visualized by the grey-colored regions with solid boundaries.



$\operatorname{supp}(v) = \overline{U_\varepsilon} = \overline{\omega}$ $\operatorname{supp}(v) = \overline{U_\varepsilon} \subseteq \omega$ $\operatorname{supp}(v) = \overline{U_\varepsilon} \subset \omega$

2.3.2 Properties of dual pairings featuring functionals of restricted support

Suppose $\lambda \in V^*$ and $v \in V$. What can we say about the dual pairing $\langle \lambda, v \rangle$ if $\operatorname{meas}(\Omega \setminus \operatorname{supp}(\lambda)) \neq \varnothing$? It is clear that the behavior of v outside of $\operatorname{supp}(\lambda)$ should have a very restricted influence on the outcome of the pairing. In this paragraph, an answer to this question is suggested which in the sequel will serve as a basis for the analysis of consistency parts of the a posteriori error estimate.

The results of this paragraph will rely on the following elementary proposition that, roughly speaking, states that for arbitrary $v \in V$ and $\omega \subseteq \Omega$ Lipschitz, the restriction $v|_\omega$ can be extended to Ω by zero. Moreover, the extension can take the parts of the boundary $\partial\omega$ where the trace of v is zero into account.

Definition 2.3.9. *Let $v \in V$ and $\omega \subseteq \Omega$ be open and Lipschitz. By $\partial\omega^0(v)$ we denote the part of the boundary $\partial\omega$ satisfying $\tau v|_{\partial\omega^0(v)} = 0$ a.e. on $\partial\omega^0(v)$ and $\tau v|_{\partial\omega \setminus \partial\omega^0(v)} \neq 0$ a.e. on $\partial\omega \setminus \partial\omega^0(v)$ simultaneously.*

Proposition 2.3.10. *Given $v \in V$ and $\omega \subseteq \Omega$ open and Lipschitz. One can always find an open Lipschitz set $\widetilde{\omega}$ with $\omega \subseteq \widetilde{\omega} \subseteq \Omega$ and a function $v_\omega^{ext} \in V_{\widetilde{\omega},0}$ with $v_\omega^{ext}|_\omega = v|_\omega$ a.e. in ω. Moreover, if $\partial\omega^0(v) \neq \varnothing$, one can always choose $\widetilde{\omega}$ so that $\partial\widetilde{\omega} \cap \partial\omega = \partial\omega^0(v)$.*

Proof. If $v \in V_{\omega,0}$, we choose $\widetilde{\omega} := \omega$. If $\overline{\omega} \subset \Omega$, a straightforward application of the extension theorem (cf. Theorem 1.3.3) proves the statement. If $\partial\omega \cap \partial\Omega \neq \varnothing$, the

extension theorem implies the existence of ω_1 open and Lipschitz satisfying $\omega \subset\subset \omega_1$ and $v_\omega^{ext} \in V_{\omega_1,0}$, without guaranteeing $\omega_1 \subset \Omega$. Requiring additionally ω_1 to be selected in such a way that the intersection $\omega_1 \cap \Omega$ is Lipschitz, we can set $\widetilde{\omega} := \omega_1 \cap \Omega$ (see the illustration in Figure 2.3). Clearly, $\partial\widetilde{\omega} = (\widetilde{\omega} \cap \partial\Omega) \cup (\partial\omega_1 \cap \overline{\Omega})$. As $\widetilde{\omega}$ is Lipschitz, the trace of v_ω^{ext} on $\partial\widetilde{\omega}$ is defined and equals to zero. Thus, $v_\omega^{ext} \in V_{\widetilde{\omega},0}$. Similarly, one can show the last statement of the proposition. □

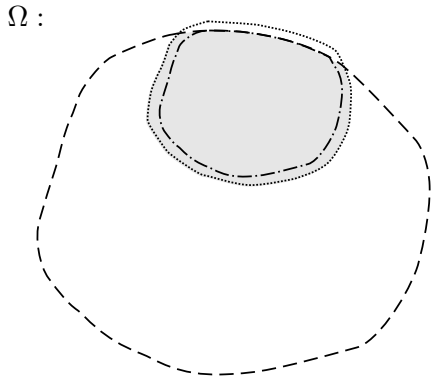


Figure 2.3: An illustration to the proof of Proposition 2.3.10. The area within the dashed line represents Ω, the area within the dash-dotted line – ω, the area within the dotted line – ω_1, the gray-colored region visualizes the set $\widetilde{\omega}$.

We can further suggest a possible way to extend the statement of Proposition 2.3.10 to the case of ω non-Lipschitz. It will require the following construction.

Definition 2.3.11. *Let $Lip(\omega)$ denote the minimal open Lipschitz set satisfying $\omega \subseteq Lip(\omega) \subseteq \Omega$.*

Obviously, the set $Lip(\omega)$ is well-defined: if no proper Lipschitz subset of Ω containing ω exists, we can take $Lip(\omega) := \Omega$ (which is Lipschitz). If ω is Lipschitz, then $Lip(\omega) := \omega$. Moreover, it seems fair to expect $\mathrm{meas}(\omega) \approx \mathrm{meas}(Lip(\omega))$. The restriction $v|_{Lip(\omega)}$ can be extended to Ω by zero in the sense of Proposition 2.3.10.

Finally, we can formulate a result that provides an interpretation of the effect V^*-functionals with restricted support on the elements of V.

Proposition 2.3.12. *For $\lambda \in V^*$, set $\Lambda := \mathrm{int}(\mathrm{supp}(\lambda))$, if $\mathrm{supp}(\lambda)$ is Lipschitz, and $\Lambda := Lip(\mathrm{int}(\mathrm{supp}(\lambda)))$, otherwise. For any $v \in V$ there exists an open Lipschitz set $\widetilde{\Lambda}$ with $\Lambda \subseteq \widetilde{\Lambda} \subseteq \Omega$, $\partial\widetilde{\Lambda} \cap \partial\Lambda = \partial\Lambda^0(v)$ and a function $v_\Lambda^{ext} \in V_{\widetilde{\Lambda},0}$ such that $v_\Lambda^{ext}|_\Lambda = v|_\Lambda$ a.e. in Λ and*

$$\langle \lambda, v \rangle = \langle \lambda, v_\Lambda^{ext} \rangle. \tag{2.3.7}$$

Proof. It suffices to prove (2.3.7), the rest follows from Proposition 2.3.10 with $\omega = \Lambda$. Let $\bar{v} := v - v_\Lambda^{ext}$. Evidently,

$$V_{\Omega\setminus\Lambda,0} \ni \bar{v} = \begin{cases} 0 & , \text{ in } \Lambda \\ v - v_\Lambda^{ext}, & \text{in } \mathrm{int}(\Omega \setminus \Lambda) \end{cases}.$$

By construction of Λ it holds $\mathrm{int}(\Omega \setminus \Lambda) \subseteq \mathcal{O}_\lambda$ (cf. Definition 2.3.7). Hence, $\langle \lambda, \bar{v} \rangle = 0$ and $\langle \lambda, v \rangle = \langle \lambda, v_\Lambda^{ext} \rangle + \langle \lambda, \bar{v} \rangle = \langle \lambda, v_\Lambda^{ext} \rangle$. □

Remark 2.3.13. It should be noted that $\langle \lambda, v \rangle = \langle \lambda, v|_{\mathrm{supp}(\lambda)} \rangle$ only if $v \in V_{int(\mathrm{supp}(\lambda)),0}$. Otherwise, λ "reaches" the values of v slightly outside of $\mathrm{int}(\mathrm{supp}(\lambda))$.

2.4 Local properties of the solution of the obstacle problem

The solution of the obstacle problem satisfies the complementarity condition

$$\psi - y \in V_+, \;\; \sigma \in V_+^*, \;\; \langle \sigma, y - \psi \rangle = 0 \tag{2.4.1}$$

(cf. (2.2.9)). To provide more accurate information on the structure of the solution (y, σ) induced by the complementarity condition, we split Ω into subsets where, in agreement with the nature of these functions, the inequalities hold either strictly or as equalities. More specifically, we aim to define the following collection of sets:

$\mathcal{A}$	:	$\{"\psi - y = 0"\}$	:	active (coincidence) set
$\mathcal{I}$	:	$\{"\psi - y > 0"\}$	:	inactive (non-coincidence) set
$\mathcal{Z}$	:	$\{"\sigma = 0"\}$	:	zero set
$\mathcal{C}$	:	$\{"\sigma > 0"\}$	:	strongly active set
$\mathcal{B}$	:	$\{"\psi - y = 0"\} \cap \{"\sigma = 0"\}$	:	biactive (weakly active) set

In addition to the knowledge of the solution's structure, these sets provide a certain convenience in the formulation of optimality systems which arise in the study of the optimal control of the obstacle problem. From the abstract definitions presented above we expect $\mathcal{A} \cup \mathcal{I} = \mathcal{Z} \cup \mathcal{C} = \Omega$, whereas, due to (2.4.1), $\mathcal{C} \subseteq \mathcal{A}$ (and, therefore, $\mathcal{I} \cup \mathcal{B} = \mathcal{Z}$) should be satisfied. We partition the domain Ω according to the definitions given in the next paragraph.

2.4.1 Definitions of sets associated with the solution of the obstacle problem

Definition 2.4.1. *(The active set, the inactive set, the free boundary w.r.t. y)*
The constraint $\psi - y \in V_+$ yields a partition

$$\Omega = \mathcal{A} \cup \mathcal{F}(y) \cup \mathcal{I},$$

where

(*i*) $\mathcal{A} := \{$ *the maximal open set D with $\psi - y = 0$ a.e. in D* $\}$ *is called the active set;*

(*ii*) *given* $B_\varepsilon(\psi - y) := \{$ *the maximal open set D where $\psi - y \geq \varepsilon$ a.e. in D* $\}$,
$\mathcal{I} := \bigcup_{\varepsilon > 0} B_\varepsilon(\psi - y)$ *is called the inactive set;*

(iii) $\mathcal{F}(y) := \Omega \setminus (\mathcal{A} \cup \mathcal{I})$ *is called the free boundary with respect to* y.

Definition 2.4.2. *(The zero set, the strongly active set, the free boundary w.r.t.* σ*)* *The constraint* $\sigma \in V_+^*$ *yields a partition*

$$\Omega = \mathcal{Z} \cup \mathcal{F}(\sigma) \cup \mathcal{C},$$

where

(i) $\mathcal{Z} := \mathcal{O}_\sigma = \{$ *the maximal open set* D *with* $\langle \sigma, v \rangle = 0, \ \forall \ v \in V_{D,0} \ \}$ *is called the zero set;*

(ii) $\mathcal{C} := int(\mathrm{supp}(\sigma))$ *(recall that* $\mathrm{supp}(\sigma) = \Omega \setminus \mathcal{O}_\sigma$*) is called the strongly active set;*

(iii) $\mathcal{F}(\sigma) := \Omega \setminus (\mathcal{Z} \cup \mathcal{C})$ *is called the free boundary with respect to* σ.

The free boundaries $\mathcal{F}(y)$ and $\mathcal{F}(\sigma)$ are typically 1-dimensional manifolds. Evidently, the supports of $\psi - y$ and σ satisfy

$$\mathrm{supp}(\psi - y) = \mathcal{I} \cup \mathcal{F}(y) \quad \text{and} \quad \mathrm{supp}(\sigma) = \mathcal{C} \cup \mathcal{F}(\sigma).$$

Remark 2.4.3. Suppose the assumptions of Lemma 2.2.6 are satisfied. We then have $y \in C(\overline{\Omega})$ and $\sigma \in L^2(\Omega)$. In this case it is natural to employ a different system of definitions for the sets, namely

$$\begin{aligned} &\mathcal{A}_{reg} := int(\{x \in \Omega \,|\, \psi(x) - y(x) = 0\}), \quad \mathcal{I}_{reg} := int(\Omega \setminus \mathcal{A}_{reg}), \\ &\mathcal{Z}_{reg} := \{ \text{ the maximal open set } D \text{ with } \sigma = 0 \text{ a.e. in } D \ \}, \quad \mathcal{C}_{reg} := int(\Omega \setminus \mathcal{Z}_{reg}), \end{aligned} \tag{2.4.2}$$

where $\mathcal{A}_{reg}$, $\mathcal{I}_{reg}$, $\mathcal{Z}_{reg}$, and $\mathcal{C}_{reg}$ are the "regular" versions of the active, inactive, zero, an strongly active sets, respectively. It should be emphasizes that Definitions 2.4.1 and 2.4.2 are universal – they can be used in both the regular setting $(y, \sigma) \in V \cap H^2(\Omega) \times L^2(\Omega)$ and the more general setting $(y, \sigma) \in V \times V^*$, whereas the system of definitions (2.4.2) is not applicable in case $(y, \sigma) \in V \times V^*$.

Finally, we can introduce the *biactive set* and the concepts of *strict complementarity* and *lack of strict complementarity* for the solution of the obstacle problem.

Definition 2.4.4. *(The biactive set, strict complementarity, lack of strict complementarity)*

(i) *The set* $\mathcal{B} := int(\mathcal{A} \setminus \mathcal{C})$ *is called the biactive set;*

(ii) *if* $\mathrm{meas}(\mathcal{B}) = 0$*, it is said that the strict complementarity condition is satisfied at the solution of the obstacle problem;*

(iii) *if* $\mathrm{meas}(\mathcal{B}) \neq 0$*, it is said that there is lack of strict complementarity at the solution of the obstacle problem.*

The strict complementarily condition requires that $\psi - y$ and σ are never zero in the same subset of Ω of a non-zero measure. It is clear that, in general, we cannot expect the strict complementarity to be satisfied. Indeed, oblivious to the shape of y, the complementarity condition is always true for $\sigma \equiv 0$.

2.4.2 Properties of the sets

The two key results of this paragraph are stated in Corollary 2.4.8 and Proposition 2.4.9, (i). The first result will support the proof of the convergence theorem (cf. Theorem 5.1.4), whereas the second one shows that the requirements to the partitions of Ω set in the beginning of this section are satisfied within the system of definitions introduced in the previous paragraph.

We start by introducing a partition of the computational domain Ω according to local properties of arbitrary functions $v \in V_+$.

Definition 2.4.5. *(The zero set, the positive set w.r.t. $v \in V_+$)*
To any $v \in V_+$, we assign

(i) *the zero set,* $\Omega^0(v) := \{$ *the maximal open set D with $v = 0$ a.e. in D* $\}$;

(ii) *the positive set,* $\Omega^+(v) := \bigcup_{\varepsilon>0} B_\varepsilon(v)$, *where*
$B_\varepsilon(v) := \{$ *the maximal open set D where $v \geq \varepsilon$ a.e. in D* $\}$.

Note that the active and the inactive set satisfy $\mathcal{I} = \Omega^+(\psi - y)$ and $\mathcal{A} = \Omega^0(\psi - y)$.

The following proposition is an intermediate step that addresses a function $v \in V_+$ and a functional $\sigma \in V_+^*$ and explains the interaction between the sets we have defined for v and σ in the case $\langle \sigma, v \rangle = 0$.

Proposition 2.4.6. *Given $\sigma \in V_+^*$ with associated sets $\mathcal{C}$ and $\mathcal{Z}$ from Definition 2.4.2. For any $v \in V_+$ satisfying $\langle \sigma, v \rangle = 0$, there holds $\Omega^+(v) \subseteq \mathcal{Z}$.*

Proof. Let $\varepsilon > 0$ be arbitrary but fixed. By the definition of $B_\varepsilon(v)$ (Definition 2.4.5, (ii)),

$$v \geq \varepsilon \text{ a.e. in } B_\varepsilon(v).$$

Since for any $\varphi \in C_0^\infty(B_\varepsilon)$ we can find $\delta(\varepsilon) > 0$ such that $\delta(\varepsilon)\varphi \leq \varepsilon/2$ a.e. in $B_\varepsilon(v)$, there holds,

$$v - \delta(\varepsilon)\varphi \geq \varepsilon/2 \text{ a.e. in } B_\varepsilon(v).$$

As $\varphi = 0$ a.e. in $\Omega \setminus B_\varepsilon$, we have $w := v - \delta(\varepsilon)\varphi \in V_+$. Now,

$$0 \leq -\underbrace{\langle \sigma, v \rangle}_{=0} + \underbrace{\langle \sigma, v - \delta(\varepsilon)\varphi \rangle}_{\geq 0} = -\delta(\varepsilon)\langle \sigma, \varphi \rangle,$$

thus, $\langle \sigma, \varphi \rangle \leq 0$. Employing the same construction with $-\varphi$, we get $\langle \sigma, \varphi \rangle \geq 0$. Since ε was chosen arbitrarily, we thus have shown

$$\langle \sigma, \varphi \rangle = 0, \ \forall\, \varphi \in C_0^\infty(B_{\varepsilon(v)}), \ \forall\, \varepsilon > 0.$$

The density of $C_0^\infty(B_\varepsilon(v))$ in $V_{B_\varepsilon(v),0}$ implies the identity

$$\langle \sigma, \varphi \rangle = 0, \ \forall\, \varphi \in V_{B_\varepsilon(v),0}, \ \forall\, \varepsilon > 0,$$

i.e., $\sigma = 0$ in $\Omega^+(v)$, and thus, $\Omega^+(v) \subseteq \mathcal{Z}$. □

Obviously, $\Omega^+(v) \subseteq \mathcal{Z}$ is equivalent to $\mathcal{C} \subseteq \Omega^0(v)$, in other words, $v = 0$ a.e. in $\mathcal{C}$.

Definition 2.4.7. *(The positive and the negative parts of $v \in L^2(\Omega)$)*
For any $v \in L^2(\Omega)$, we set

$$v^+ := \max(0, v), \quad v^- := \min(0, v),$$

where the operators "max" and "min" are to be understood in the pointwise a.e. sense.

Corollary 2.4.8. *Let σ and $\mathcal{C}$ be as in Proposition 2.4.6. For any $v \in V$ with $\langle \sigma, v^+ \rangle = \langle \sigma, v^- \rangle = 0$ there holds $v = 0$ a.e. in $\mathcal{C}$, $\langle \sigma, v \rangle = 0$.*

Proof. Due to $\langle \sigma, v^+ \rangle = \langle \sigma, v^- \rangle = 0$ we have $v^+ = v^- = 0$ a.e. in $\mathcal{C}$. As $v = v^+ + v^-$, the assertion follows. □

We further observe the properties of σ in the set $\mathcal{A}$.

Proposition 2.4.9. *The slack variable σ satisfies*

(*i*) *$\sigma = 0$ in $\mathcal{I}$, i.e., $\mathcal{C} \subseteq \mathcal{A}$;*

(*ii*) *$\sigma = g + \Delta\psi$ in $\mathcal{A}$.*

Proof. (*i*): Direct application of Proposition 2.4.6 with $v = \psi - y$. (*ii*): The functional $\Delta\psi \in V^*$ is defined as

$$\langle \Delta\psi, v \rangle := -(\nabla\psi, \nabla v)_{0,\Omega}, \ \forall\, v \in V.$$

Then, in view of (2.2.9a) and (2.2.3), for any $v \in V_{\mathcal{A},0}$,

$$\begin{aligned} \langle \sigma, v \rangle &= (g, v)_{0,\Omega} - (\nabla\psi, \nabla v)_{0,\mathcal{A}} = (g, v)_{0,\Omega} - (\nabla\psi, \nabla v)_{0,\Omega} = \\ &= (g, v)_{0,\Omega} + \langle \Delta\psi, v \rangle = \langle g + \Delta\psi, v \rangle, \end{aligned}$$

which proves the assertion. □

Corollary 2.4.10. *Lack of strict complementarity at the solution (y, σ) is observed if and only if there is a set $\mathcal{B} \subseteq \mathcal{A}$ such that $g + \Delta\psi = 0$ in $\mathcal{B}$. This can happen only if*

$$\langle \Delta\psi, v \rangle = -(g, v)_{0,\mathcal{B}}, \ \forall\, v \in V_{\mathcal{B},0}.$$

In other words, we must have $\Delta\psi|_{\mathcal{B}} \in L^2(\mathcal{B})$.

3 Optimal control of the obstacle problem

3.1 The optimal control problem

We consider the following distributed optimal control problem where the obstacle problem given in the form of the variational inequality (2.2.7) determines the set of constraints:

$$\begin{aligned} \textit{minimize} \quad & J(y,u) := \frac{1}{2}\,\|y-y^d\|^2_{0,\Omega} \;+\; \frac{\alpha}{2}\,\|u-u^d\|^2_{0,\Omega} && \text{(3.1.1a)}\\ \textit{over} \quad & (y,u) \in V \times L^2(\Omega), \\ \textit{subject to} \quad & a(y, y-v) \;\le\; (f+u, y-v)_{0,\Omega}\,, \quad \forall\, v \in K, && \text{(3.1.1b)}\\ & y \in K = \{v \in V \mid v \le \psi \text{ a.e. in } \Omega\}, \end{aligned}$$

where it is assumed that

$$y^d \in L^2(\Omega),\; u^d \in L^2(\Omega),\; f \in L^2(\Omega),\; \psi \in V,\; \alpha \in \mathbb{R}_+. \tag{3.1.2}$$

Here J is referred to as the *objective functional*. The data functions ψ, f, y^d, and u^d stand for the *obstacle*, the *force term*, the *desired state*, and the *shift control*, respectively. The system is controlled via the additional force term $u \in L^2(\Omega)$ - the *control function*, - distributed in the whole computational domain Ω (thus a distributed control problem). We refer to y as the *state function*.
The optimal control problem (3.1.1) can be equivalently written in the so-called control-reduced form, eliminating the state y by means of the control-to-state mapping (cf. Definition 2.1.1):

$$\begin{aligned} \textit{minimize} \quad & J^{red}(u) := \frac{1}{2}\,\|Su-y^d\|^2_{0,\Omega} \;+\; \frac{\alpha}{2}\,\|u-u^d\|^2_{0,\Omega} && \text{(3.1.3)}\\ \textit{over} \quad & u \in L^2(\Omega). \end{aligned}$$

The existence of minimizers for (3.1.1) is guaranteed by the following theorem:

Theorem 3.1.1. *(Existence of an optimal solution)*
If the assumptions (3.1.2) *on the data are satisfied, the optimal control problem* (3.1.1) *admits an optimal solution* $(y,u) \in V \times L^2(\Omega)$*. In general, there is no uniqueness.*

The proof of the theorem is given in [49]. For the sake of completeness, it is included below.

Proof. Obviously, there holds

$$J^{red} : L^2(\Omega) \to \mathbb{R}, \quad J^{red}(u) \geq 0, \forall\, u \in L^2(\Omega).$$

Thus, by completeness of $\mathbb{R}$, there exists $\mathbb{R} \ni j := \inf_{u \in L^2(\Omega)} J^{red}(u)$. Let $\{u_n\}_{n\in\mathbb{N}} \subset L^2(\Omega)$ be a minimizing sequence for (3.1.3), i.e.,

$$\lim_{n\to\infty} J^{red}(u_n) = j.$$

By Young's inequality with $\gamma = 1/2$ we have

$$J^{red}(u) \geq \frac{\alpha}{2}\|u - u^d\|^2_{0,\Omega} \geq \frac{\alpha}{4}\|u\|^2_{0,\Omega} - \frac{\alpha}{2}\|u^d\|^2_{0,\Omega}.$$

As α is strictly positive, we have $\|u_n\|^2_{0,\Omega} \leq 4\alpha^{-1}J^{red}(u_n) + 2\|u^d\|^2_{0,\Omega}$, $\forall\, n \in \mathbb{N}$, i.e., the minimizing sequence has to be bounded in $L^2(\Omega)$. Since $L^2(\Omega)$ is a reflexive, separable Banach space, every bounded sequence in $L^2(\Omega)$ admits a weakly convergent subsequence. Therefore, there exist $u^* \in L^2(\Omega)$ and a subsequence $\{u_{n_k}\}_{k\in\mathbb{N}}$ such that

$$u_{n_k} \rightharpoonup u^* \text{ in } L^2(\Omega).$$

The next step is to show that u^* is an optimal solution. The embedding of $L^2(\Omega)$ into V^* is compact and so

$$u_{n_k} \to u^* \text{ in } V^*.$$

Consequently, due to the continuity of the control-to-state mapping,

$$Su_{n_k} \to Su^* \text{ in } V.$$

Thus,

$$\liminf_{k\to\infty} J^{red}(u_{n_k}) = \frac{1}{2}\,\|Su^* - y^d\|^2_{0,\Omega} + \frac{\alpha}{2}\liminf_{k\to\infty}\|u_{n_k} - u^d\|^2_{0,\Omega}.$$

The functional

$$\begin{aligned} G: \quad L^2(\Omega) &\longrightarrow \mathbb{R}, \\ u &\longmapsto \|u - u^d\|^2_{0,\Omega} \end{aligned}$$

is lower semicontinuous and convex and thus weakly lower semicontinuous (see, e.g., [51], Proposition 5.4.4). Therefore,

$$J^{red}(u^*) \leq \liminf_{k\to\infty} J^{red}(u_{n_k}) \leq j,$$

i.e., u^* is an optimal solution. Due to the fact that the underlying variational inequality is uniquely solvable, each optimal control u^* gives rise to a unique state $y^* := Su^*$. However, with the reference to Section 2.2 we know that the control-to-state operator S is non-convex, hence, the reduced objective functional is non-convex and we cannot guarantee the uniqueness of the minimizer u^*. □

3.2 Alternative formulations

The optimal control problem (3.1.1) is an example of a so-called Mathematical Problem with Equilibrium Constraints (MPEC) in functional space. Commonly, optimal control problems with variational inequalities in the set of constraints are referred to as MPEC.

Depending on what serves our purposes best, any of the equivalent formulations of the variational inequality (3.1.1b) enlisted in the previous chapter can be used to describe the set of constraints of (3.1.1). For instance, we can rewrite the constraints in the form of the complementarity system (2.2.9), i.e.,

$$\begin{aligned} \textit{minimize} \quad & J(y,u) := \frac{1}{2}\,\|y-y^d\|^2_{0,\Omega} \;+\; \frac{\alpha}{2}\,\|u-u^d\|^2_{0,\Omega} && \text{(3.2.1a)}\\ \textit{over} \quad & (y,\sigma,u) \in V \times V^* \times L^2(\Omega), \\ \textit{subject to} \quad & a(y,v) = (f+u,v)_{0,\Omega} - \langle \sigma, v\rangle, \quad \forall\, v \in V, && \text{(3.2.1b)}\\ & \psi - y \in V_+,\ \sigma \in V_+^*,\ \langle \sigma, y-\psi\rangle = 0. && \text{(3.2.1c)} \end{aligned}$$

The problem (3.2.1) is commonly referred to as a Mathematical Program with Complementarity Constraints (MPCC). MPCC is a subclass of MPEC.

From now on, we will often work with the formulation (3.2.1) of the optimal control problem instead of the original formulation given in (3.1.1).

3.3 Necessary optimality conditions

In this section, various concepts of stationarity associated with the optimal control problem (3.1.1) are presented and discussed. Stationary points are solutions of systems of first order (necessary) optimality conditions. For the results related to second order optimality conditions for the problem (3.1.1), the reader is referred to [46].

A classical method for the derivation of first order optimality conditions relies on the total Lagrange functional for the MPCC problem (3.2.1),

$$\mathcal{L}(y,u,\sigma,\underline{p},\underline{\mu},\underline{q},\underline{\kappa}) = J(y,u) - \langle Ay+\sigma-u-f,\underline{p}\rangle - \langle \underline{\mu}, \psi - y\rangle - \langle \sigma, \underline{q}\rangle - \underline{\kappa}\langle \sigma, \psi - y\rangle,$$

and requires the verification of the existence of Lagrange multipliers $(p,\mu.q,\kappa) \in V \times V^* \times V \times \mathbb{R}$ (underlined in the above expression). The condition for the existence of the multipliers is the fulfillment of so-called constraint qualifications (CQ). Whenever some CQ is satisfied for an optimal control $u \in L^2(\Omega)$, the multipliers exist and, together with the corresponding functions $(y,\sigma) \in V \times V^*$ and the control function u, they fulfill a KKT type system of necessary optimality conditions. This scheme experiences difficulties when applied to the problem (3.2.1). The source of the difficulties is the complementarity constraint (3.2.1c). Indeed, as we observed in Paragraph 2.2.2, the interior of the admissible set $\mathcal{D}$ corresponding to the complementarity constraint is empty, so there is no feasible solution that satisfies all the inequalities strictly. Hence, all the classical constraint qualifications (such as, for instance, the Mangasarian-Fromovitz

constraint qualification (MFCQ), cf. [56]) are violated at every feasible point of (3.2.1). Thus, it is not possible to derive optimality conditions with the help of the standard mathematical programming theory in Banach spaces.

An in-depth study of the question of the existence of the Lagrange multipliers for the problem (3.2.1) is performed in [10]. There, counter examples are presented that show that even if the strict complementarity condition is satisfied on the optimal solution of (3.2.1), it may happen that the full Lagrange system induced by $\mathcal{L}$ has no solution.

Remark 3.3.1. It should be noted that such constraint qualifications as the Slater condition cannot be applied to (3.2.1) as the control-to-state mapping is non-convex (cf. Proposition 2.2.3). (the Slater assumption plays a key role in the derivation of the necessary optimality system for a somewhat related problem of optimal control of elliptic partial differential equations with state constraints).

In the literature, the approaches that have been considered to obtain optimality systems for (3.1.1) can roughly be divided into two groups: those based on convex analysis techniques (see, e.g., [49]) and those relying on approximation methods (see, e.g., [5] or [41]). In either instance, the main concern is to obtain a system of conditions that describes optimal solutions as complete as possible. In the following sections, optimality systems relevant for this thesis will be introduced.

Remark 3.3.2. The problem under consideration is non-convex, hence, any solution of a first order optimality system is only a *stationary point* that does not need to be the minimizer of the objective functional.

3.3.1 Case $(y, \sigma) \in V \cap H^2(\Omega) \times L^2(\Omega)$

Already for MPCC in finite dimensions, derivation of optimality conditions is challenging. Various approaches result in a hierarchy of stationarity concepts, such as B(ouligand)-, W(eak)-, C(larke)-, M(ordukhovich)-, and S(trong)-stationarity. This paragraph gives a brief overview of the generalization of the concepts of C- and S-stationarity for MPEC in Hilbert space that has been offered in [33].

Let the stronger set of assumptions of Lemma 2.2.6 be satisfied. Problem (3.2.1) can then be reformulated as

$$\begin{aligned}
&\textit{minimize} && J(y,u) := \frac{1}{2}\,\|y - y^d\|^2_{0,\Omega} \;+\; \frac{\alpha}{2}\,\|u - u^d\|^2_{0,\Omega} && \text{(3.3.1a)}\\
&\textit{over} && (y,\sigma,u) \in V \times L^2(\Omega) \times L^2(\Omega),\\
&\textit{subject to} && a(y,v) = (f + u - \sigma, v)_{0,\Omega}\,,\ \forall\, v \in V, && \text{(3.3.1b)}\\
& && \psi - y \geq 0 \text{ a.e. in } \Omega,\ \sigma \geq 0 \text{ a.e. in } \Omega,\ (\sigma, y - \psi)_{0,\Omega} = 0. && \text{(3.3.1c)}
\end{aligned}$$

Now, following [33], we introduce various stationarity concepts for problem (3.3.1), employing three different concepts of localization of V^*-functionals that has been addressed in Remark 2.3.8.

Definition 3.3.3. *($\varepsilon-$almost (C-) S-stationary, almost (C-) S-stationary, (C-) S-stationary points of* (3.3.1)*)*
Let $(y, \sigma, u, p, \mu) \in V \times L^2(\Omega) \times L^2(\Omega) \times V \times V^*$ be given and satisfy

$$a(y, v) = (f + u - \sigma, v)_{0,\Omega}\,, \ \forall\, v \in V, \tag{3.3.2a}$$
$$\psi - y \geq 0 \ \text{a.e. in } \Omega, \ \sigma \geq 0 \ \text{a.e. in } \Omega, \ (\sigma, y - \psi)_{0,\Omega} = 0, \tag{3.3.2b}$$
$$a(p, v) = (y^d - y, v)_{0,\Omega} - \langle \mu, v \rangle\,, \ \forall\, v \in V, \tag{3.3.2c}$$
$$p = \alpha(u - u^d), \tag{3.3.2d}$$
$$p = 0 \ \text{a.e. in } \mathcal{C}, \tag{3.3.2e}$$
$$\langle \mu, p \rangle \geq 0, \tag{3.3.2f}$$
$$\langle \mu, y - \psi \rangle = 0. \tag{3.3.2g}$$

($i.a$) The triple $(y, \sigma, u) \in V \times L^2(\Omega) \times L^2(\Omega)$ is called an $\varepsilon-$almost C-stationary point of (3.3.1) *if the pair $(p, \mu) \in V \times V^*$ fulfills*

$$\forall \varepsilon > 0 \ \exists\, U_\varepsilon \subseteq \mathcal{I} \text{ with } \operatorname{meas}(\mathcal{I} \backslash U_\varepsilon) \leq \varepsilon \text{ such that } \langle \mu, v \rangle = 0, \ \forall\, v \in V_{U_\varepsilon}. \tag{3.3.3}$$

($i.b$) The triple $(y, \sigma, u) \in V \times L^2(\Omega) \times L^2(\Omega)$ is called an almost C-stationary point of (3.3.1) *if the pair $(p, \mu) \in V \times V^*$ fulfills*

$$\langle \mu, v \rangle = 0, \ \forall\, v \in V_{\mathcal{I},0}\,. \tag{3.3.4}$$

($i.c$) The triple $(y, \sigma, u) \in V \times L^2(\Omega) \times L^2(\Omega)$ is called a C-stationary point of (3.3.1) *if the pair $(p, \mu) \in V \times V^*$ fulfills*

$$\langle \mu, v \rangle = 0, \ \forall\, v \in V_{\mathcal{I}}\,. \tag{3.3.5}$$

(ii) Let $(y, \sigma, u) \in V \times L^2(\Omega) \times L^2(\Omega)$ be an $\varepsilon-$almost C-stationary, an almost C-stationary, or a C-stationary point of (3.3.1)*. We call (y, σ, u) an $\varepsilon-$almost S-stationary, an almost S-stationary, or a S-stationary point of* (3.3.1)*, respectively, if the corresponding pair $(p, \mu) \in V \times V^*$ additionally satisfies*

$$\langle \mu, v \rangle \geq 0, \ \forall\, v \in V_{\mathcal{B}} \cap V_+\,, \tag{3.3.6a}$$
$$p \geq 0 \ \text{a.e. in } \mathcal{B}. \tag{3.3.6b}$$

In what follows, we will refer to p as *the adjoint state*, to equation (3.3.2c) as *the adjoint equation*, and to the functional μ as *the Lagrangian multiplier associated with the adjoint state equation.*

In the latter definition, S-stationarity is the strongest concept, whereas $\varepsilon-$almost C-stationarity is the weakest one. The hierarchy of all the concepts introduced above is shown in the scheme below.

S-stationarity	$\Rightarrow$	almost S-stationarity	$\Rightarrow$	ε-almost S-stationarity
$\Downarrow$		$\Downarrow$		$\Downarrow$
C-stationarity	$\Rightarrow$	almost C-stationarity	$\Rightarrow$	ε-almost C-stationarity

Note that in [33] the names for the above stationarity concepts were chosen in account of the similarities to C- and S-stationarity concepts in the finite-dimensional case.

Remark 3.3.4. In [33], in the definitions of $\varepsilon-$almost, almost S-stationarity, and S-stationarity concepts, condition (3.3.6a) is combined with (3.3.3), (3.3.4), and (3.3.5), resulting in

$$\langle \mu, v \rangle \geq 0, \ \forall\, v \in V_{\mathcal{B} \cup U_\varepsilon}\,, \ v \geq 0 \text{ a.e. in } \mathcal{B}, \tag{3.3.7}$$

for $\varepsilon-$almost S-stationarity,

$$\langle \mu, v \rangle \geq 0, \ \forall\, v \in V_{\mathcal{Z}}\,, \ v \geq 0 \text{ a.e. in } \mathcal{B}, \ \max(0, -v)|_{\mathcal{I}} \in H_0^1(\mathcal{I}), \tag{3.3.8}$$

for almost S-stationarity, and

$$\langle \mu, v \rangle \geq 0, \ \forall\, v \in V_{\mathcal{Z}}\,, \ v \geq 0 \text{ a.e. in } \mathcal{B}, \tag{3.3.9}$$

for S-stationarity, respectively. It is easy to see that the pair of conditions (3.3.6a), (3.3.3) is equivalent to (3.3.7). Indeed, as $\mathcal{B} \cap U_\varepsilon = \varnothing$, any v from (3.3.7) can be written as $v = v_1 + v_2$, where v_1 satisfies the conditions of (3.3.6a) and v_2 of (3.3.3); the other direction can be shown by testing (3.3.7) with v, $-v \in V_{U_\varepsilon}$ and $v \in V_{\mathcal{B}}$ admissible in (3.3.7). The validity of the expressions (3.3.8) and (3.3.9) can be shown in a similar way.

Remark 3.3.5. In the definitions of the "stronger" concepts some conditions duplicate each other. Namely, condition (3.3.5) automatically implies (3.3.2g), whereas conditions (3.3.6) naturally lead to (3.3.2f).

3.3.2 Case $(y, \sigma) \in V \times V^*$

In [49], the authors derive the following system of first order necessary optimality conditions using the concept of directional (conical) derivative for the control-to-state map.

Theorem 3.3.6. *Let $(y, \sigma, u) \in V \times V^* \times L^2(\Omega)$ be an optimal solution of (3.2.1). Then there exists a pair $(p, \mu) \in V \times V^*$ that together with the triple (y, σ, u) satisfies the following conditions:*

$$p \geq 0 \text{ a.e. in } \mathcal{A}, \langle \sigma, p \rangle = 0, \tag{3.3.10a}$$

$$a(p, v) = (y^d - y, v)_{0,\Omega} - \langle \mu, v \rangle\,, \quad \forall\, v \in V, \tag{3.3.10b}$$

$$\langle \mu, v \rangle \geq 0\,, \quad \forall\, v \in V \text{ such that } v \geq 0 \text{ a.e. in } \mathcal{A}, \ \langle \sigma, v \rangle = 0, \tag{3.3.10c}$$

$$p = \alpha(u - u^d). \tag{3.3.10d}$$

In the optimality system (3.3.10) the sets $\mathcal{B}$ and $\mathcal{C}$ are not mentioned explicitly. In some situations, such formulation can be favorable.

It is easy to see that, in case $(y, \sigma) \in V \cap H^2(\Omega) \times L^2(\Omega)$, the set of conditions (3.2.1), (3.3.10) is identical to the system for the S-stationarity concept from Definition 3.3.3. Therefore, we will refer to the solutions of system (3.2.1), (3.3.10) as the *S-stationary points of* (3.1.1).

Remark 3.3.7. As there is no uniqueness for the original optimal control problem (3.1.1), there is no uniqueness for the solutions of the S-stationarity system introduced above. In fact, there are at least as many S-stationary points of one type as there are minimizers of (3.1.1).

Motivated by the convergence result that will be presented in Chapter 5 and by the analogy to the hierarchy of Definition 3.3.3, we define the following set of C-stationary concepts for problem (3.1.1).

Definition 3.3.8. *($\varepsilon-$almost C-stationary, almost C-stationary, C-stationarity points of* (3.1.1)*)*
Let $(y, \sigma, u, p, \mu) \in V \times V^* \times L^2(\Omega) \times V \times V^*$ be given and satisfy

$$a(y, v) = (f + u, v)_{0,\Omega} - \langle \sigma, v \rangle \,, \ \forall\, v \in V, \tag{3.3.11a}$$
$$\psi - y \in V_+, \ \sigma \in V_+^*, \ \langle \sigma, \psi - y \rangle = 0, \tag{3.3.11b}$$
$$a(p, v) = (y^d - y, v)_{0,\Omega} - \langle \mu, v \rangle \,, \ \forall\, v \in V, \tag{3.3.11c}$$
$$p = \alpha(u - u^d), \tag{3.3.11d}$$
$$p = 0 \ \textit{a.e. in } \mathcal{C}, \tag{3.3.11e}$$
$$\langle \mu, p \rangle \geq 0, \tag{3.3.11f}$$
$$\langle \mu, y - \psi \rangle = 0. \tag{3.3.11g}$$

(i) The triple $(y, \sigma, u) \in V \times V^* \times L^2(\Omega)$ is called an $\varepsilon-$almost C-stationary point of (3.1.1) *if the pair $(p, \mu) \in V \times V^*$ fulfills*

$$\forall \varepsilon > 0 \ \exists\, U_\varepsilon \subseteq \mathcal{I} \textit{ with } \operatorname{meas}(\mathcal{I} \backslash U_\varepsilon) \leq \varepsilon \textit{ such that } \langle \mu, v \rangle = 0, \ \forall\, v \in V_{U_\varepsilon}. \tag{3.3.12}$$

(ii) The triple $(y, \sigma, u) \in V \times V^* \times L^2(\Omega)$ is called an almost C-stationary point of (3.1.1) *if the pair $(p, \mu) \in V \times V^*$ fulfills*

$$\langle \mu, v \rangle = 0, \ \forall\, v \in V_{\mathcal{I},0} \,. \tag{3.3.13}$$

(iii) The triple $(y, \sigma, u) \in V \times V^* \times L^2(\Omega)$ is called a C-stationary point of (3.1.1) *if the pair $(p, \mu) \in V \times V^*$ fulfills*

$$\langle \mu, v \rangle = 0, \ \forall\, v \in V_{\mathcal{I}} \,. \tag{3.3.14}$$

Remark 3.3.9.

(*i*) Assuming additionally $\psi \in L^\infty(\Omega) \cap V$, one could use the regularity of the state function $y \in L^\infty(\Omega) \cap V$ to obtain $\mu \in \mathcal{M}(\overline{\Omega})$ instead of $\mu \in V^*$ (cf. Lemma 2.2.5).

(*ii*) If a system of necessary conditions without explicit mention of $\mathcal{C}$ is desired, condition (3.3.11e) can be replaced with $\langle \sigma, p^+ \rangle = \langle \sigma, p^- \rangle = 0$.

(*iii*) Comparing to the discrete C-stationarity system (see Definition 4.4.8 and Remark 4.4.9), it seems appropriate to complete all the above versions of C-stationarity concepts with the condition

$$\langle \mu, pv \rangle \geq 0, \quad \forall\, v \in C^1(\overline{\Omega}) \cap (V_{\mathcal{B},0})_+ \quad (\text{or } v \in C^1(\overline{\Omega}) \cap (V_{\mathcal{B}})_+).$$

Remark 3.3.10. If one tries to formally derive the optimality system using the Lagrangian $\mathcal{L}$ (cf. Section 3.3), the functional μ plays the role of a multiplier for the state constraint $\psi - y \in V_+$, the adjoint state p can be identified with the multiplier $-q$ and thus, plays the role of a multiplier for the constraint $\sigma \in V_+^*$ and the state equation (3.2.1b) simultaneously. The constraint $\langle \sigma, \psi - y \rangle = 0$ is not treated explicitly as the existence of the corresponding Lagrange multiplier κ does not follow from the first order optimality conditions (cf. [10]).

3.3.3 On the local structure of stationary points

In this paragraph, some of the local properties of almost C-stationary points with respect to the sets defined in Section 2.4 are listed. An overview of the properties is given in the table below followed by a more detailed explanation.

	$\mathcal{C}$	$\mathcal{B}$	$\mathcal{I}$
y	$= \psi$ a.e.	$= \psi$ a.e.	–
p	$= 0$ a.e.	$= -\alpha(\Delta\psi + f + u^d)$ a.e.	–
u	$= u^d$ a.e.	$= -\Delta\psi - f$ a.e.	–
σ	$= f + u^d + \Delta\psi$	$= 0$	$= 0$
μ	$= y^d - \psi$	$= y^d - \psi + \alpha\Delta(\Delta\psi + f + u^d)$	$= 0$

By the definitions of $\mathcal{A}$, $\mathcal{C}$, and $\mathcal{B}$, there holds $y = \psi$ a.e. in $\mathcal{A} = \mathcal{C} \cup \mathcal{B}$. As $V_{\mathcal{B},0} \subseteq V_{\mathcal{Z},0}$ and $V_{\mathcal{I},0} \subseteq V_{\mathcal{Z},0}$, we obviously have

$$\langle \sigma, v \rangle = 0, \quad \forall\, v \in V_{\mathcal{B},0}, \quad \langle \sigma, v \rangle = 0, \quad \forall\, v \in V_{\mathcal{I},0}.$$

In $\mathcal{C}$, due to (3.3.11d) and (3.3.11e),

$$p = 0 \text{ a.e. in } \mathcal{C}, \quad u = u^d \text{ a.e. in } \mathcal{C}.$$

Thus, in particular, (3.3.11c) implies

$$\langle \mu, v \rangle = (y^d - \psi, v)_{0,\mathcal{C}}, \ \forall\ v \in V_{\mathcal{C},0},$$

i.e., the restriction of μ to $\mathcal{C}$ can be identified with $\mu_{\mathcal{C}} := y^d - \psi \in L^2(\mathcal{C})$.
In view of (2.2.3) and by (3.3.11a),

$$\langle \sigma, v \rangle = (f + u^d, v)_{0,\mathcal{C}} - (\nabla\psi, \nabla v)_{0,\mathcal{C}}, \ \forall\ v \in V_{\mathcal{C},0},$$

thus, $\sigma = f + u^d + \Delta\psi$ in $\mathcal{C}$ (see Proposition 2.4.9 for the definition of $\Delta\psi \in V^*$).
In $\mathcal{B}$, we have

$$(f + u, v)_{0,\mathcal{B}} = (\nabla\psi, \nabla v)_{0,\mathcal{B}}, \ \forall\ v \in V_{\mathcal{B},0}.$$

This means that the weak divergence of $\nabla\psi$ in $\mathcal{B}$ exists and is equal to $-(f + u)|_{\mathcal{B}} \in L^2(\mathcal{B})$, thus, $-\Delta\psi = f + u$ a.e. in $\mathcal{B}$ and $\Delta\psi \in L^2(\mathcal{B})$. Therefore,

$$u = -\Delta\psi - f \text{ a.e. in } \mathcal{B},$$

and, due to (3.3.11d),

$$p = -\alpha(\Delta\psi + f + u^d) \text{ a.e. in } \mathcal{B};$$

Note that the latter equivalence implies $\Delta\psi + f + u^d \in H^1(\mathcal{B})$. Hence, by (3.3.11c), without any additional regularity assumptions on the data functions, we obtain

$$\langle \mu, v \rangle = (y^d - \psi, v)_{0,\mathcal{B}} - \alpha\,(\nabla(\Delta\psi + f + u^d), \nabla v)_{0,\mathcal{B}}, \ \forall\ v \in V_{\mathcal{B},0}.$$

The latter can be compactly written as

$$\mu = y^d - \psi + \alpha\Delta(\Delta\psi + f + u^d) \text{ in } \mathcal{B},$$

where

$$(V_{\mathcal{B},0})^* \ni \Delta(\Delta\psi + f + u^d) := (\nabla(\Delta\psi + f + u^d), \nabla v)_{0,\mathcal{B}}, \ \forall\ v \in V_{\mathcal{B},0}.$$

4 Discrete optimal control problem

In order to make a numerical treatment of the optimal control problem (3.1.1) possible, a discretization is necessary. In this chapter, an introduction will be given for an approximation scheme for the optimal control problem and for discrete stationarity concepts. Also, some further objects related to the discrete solutions are defined that will be used in later chapters.

4.1 Preliminaries

4.1.1 Discretization of the domain

By $\mathcal{T}_\ell$ we denote a shape-regular simplicial conforming triangulation of $\overline{\Omega}$ which aligns with Γ. Let $\mathcal{N}_\ell(D)$, $\mathcal{E}_\ell(D)$ and $\mathcal{T}_\ell(D)$ denote the sets of nodes, edges and triangles of $\mathcal{T}_\ell$ in $D \subseteq \overline{\Omega}$. Further, we distinguish the sets of overall nodes and edges

$$\overline{\mathcal{N}}_\ell := \mathcal{N}_\ell(\overline{\Omega}), \quad \overline{\mathcal{E}}_\ell := \mathcal{E}_\ell(\overline{\Omega}),$$

the sets of boundary nodes and boundary edges

$$\mathcal{N}_\ell^\Gamma := \mathcal{N}_\ell(\Gamma), \quad \mathcal{E}_\ell^\Gamma := \mathcal{E}_\ell(\Gamma),$$

and the sets of interior nodes and edges

$$\mathcal{N}_\ell := \overline{\mathcal{N}}_\ell \setminus \mathcal{N}_\ell^\Gamma, \quad \mathcal{E}_\ell := \overline{\mathcal{E}}_\ell \setminus \mathcal{E}_\ell^\Gamma.$$

We refer to h_T and $|T|$ as the diameter and the area of an element $T \in \mathcal{T}_\ell$, respectively, whereas h_E stands for the length of an edge $E \in \mathcal{E}_\ell$.

For every $E \in \mathcal{E}_\ell$, one can find $T_\pm \in \mathcal{T}_\ell$ such that $E = T_+ \cap T_-$. Let ν_E be the unit normal to E directed towards T_- and set $\omega_E := T_+ \cup T_-$. For $a \in \overline{\mathcal{N}}_\ell$, $E \in \mathcal{E}_\ell$, and $T \in \mathcal{T}_\ell$ we refer to

$$\begin{aligned}
\omega_\ell^a &:= \bigcup\{T \in \mathcal{T}_\ell \mid a \in \mathcal{N}_\ell(T)\},\\
\omega_\ell^E &:= \bigcup\{T \in \mathcal{T}_\ell \mid E \in \mathcal{E}_\ell(T)\},\\
\omega_\ell^T &:= \bigcup\{T' \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T') \cap \mathcal{N}_\ell(T) \neq \varnothing\}
\end{aligned}$$

as the patches of triangles having a nonempty intersection with the vertex a, the edge E, and the triangle T, respectively. For $a \in \mathcal{N}_\ell$, we further define

$$\mathcal{E}_\ell^a := \bigcup\{E \in \mathcal{E}_\ell \mid a \in \mathcal{N}_\ell(E)\}$$

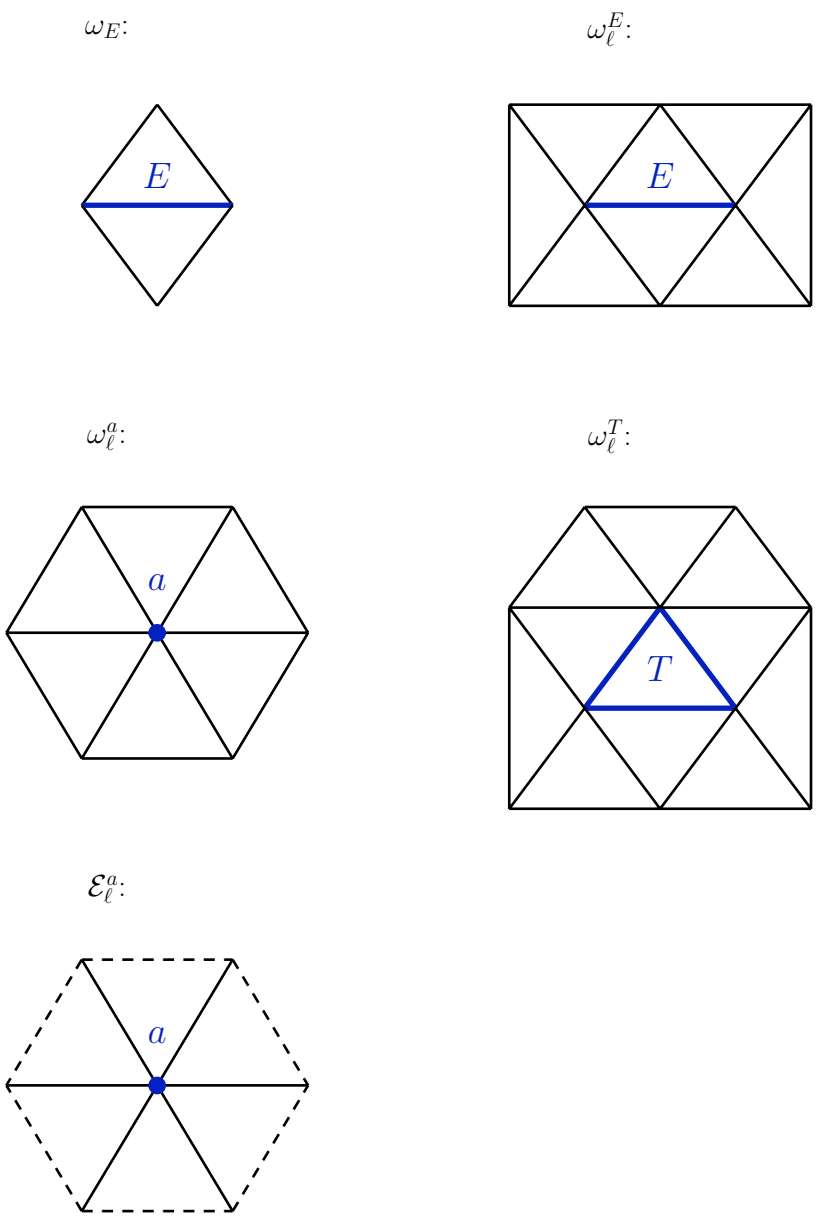


Figure 4.1: From top left to the bottom: patches of triangles associated with the edge E (ω_E and ω_ℓ^E), the nodal point a (ω_ℓ^a), the triangle T (ω_ℓ^T) and a patch of edges $\mathcal{E}_\ell^a$ associated with the nodal point a.

as the set of edges sharing a as a common vertex (see Figure 4.1 for the illustration of the patches).

Finally, for comparison of two quantities $A,\ B \in \mathbb{R}$ we will use the notation $A \lesssim B$, meaning that there exists a positive constant C, independent on the local mesh size, such that $A \leq CB$. The constant C may only depend on the shape regularity of the triangulation $\mathcal{T}_\ell$, the regularization parameter α and the coercivity constant γ_a.

4.1.2 Discretization of the functional spaces

Referring to $P_k(T)$, $T \in \mathcal{T}_\ell$, $k \in \mathbb{N}_0$, as the linear space of polynomials of degree k on T, we denote by

$$W_\ell := \{v_\ell \in L^2(\Omega) \mid v_\ell|_T \in P_0(T)\,,\ T \in \mathcal{T}_\ell\}$$

the finite element space of piecewise constant functions and by

$$S_\ell := \{v_\ell \in C(\overline{\Omega}) \mid v_\ell|_T \in P_1(T)\,,\ T \in \mathcal{T}_\ell\}$$

the finite element space of continuous piecewise affine functions. The latter space spans the canonical nodal basis functions associated with the nodal points of $\mathcal{T}_\ell$: $\varphi_\ell^{(a)}$, $a \in \overline{\mathcal{N}}_\ell$. By V_ℓ we denote the subspace of S_ℓ with functions vanishing on the boundary of Ω,

$$V_\ell := \{v_\ell \in S_\ell \mid v_\ell|_\Gamma = 0\} = \operatorname{span}\{\varphi_\ell^{(a)},\ a \in \mathcal{N}_\ell\}.$$

For every $v_\ell \in V_\ell$, on every interior edge $E \in \mathcal{E}_\ell$ we can define the *jump* of the weak gradient $\nabla v_\ell \in L^2(\Omega;\mathbb{R}^2)$ according to

$$\nu_E \cdot [\nabla v_\ell] := \nu_E \cdot (\nabla v_\ell|_{T_+} - \nabla v_\ell|_{T_-}).$$

The number of the interior nodes $N_\ell := \operatorname{card}(\mathcal{N}_\ell)$ gives the number of the degrees of freedom in V_ℓ. As the dual space of V_ℓ, we consider linear combinations of the Dirac delta functionals δ_a associated with $a \in \mathcal{N}_\ell$, i.e.,

$$\mathcal{M}_\ell \ :=\ \{\lambda_\ell \in \mathcal{M}(\overline{\Omega}) \mid \lambda_\ell = \sum_{a \in \mathcal{N}_\ell} \lambda_\ell(a)\delta_a\ ,\ \lambda_\ell(a) \in \mathbb{R}\}. \tag{4.1.1}$$

The real numbers $\lambda_\ell(a),\ a \in \mathcal{N}_\ell$ are the weights characterizing the distribution $\lambda_\ell \in \mathcal{M}_\ell$. We have $\dim \mathcal{M}_\ell = \dim V_\ell = N_\ell$.

Remark 4.1.1. One should emphasize that $\mathcal{M}_\ell$ spans the delta functionals associated with the interior nodal points only. For $a \in \mathcal{N}_\ell$, the coefficients of λ_ℓ satisfy

$$\lambda_\ell(a) = \langle\langle \lambda_\ell, \varphi_\ell^{(a)} \rangle\rangle, \quad \forall a \in \mathcal{N}_\ell. \tag{4.1.2}$$

4.1.3 $S_\ell-$approximation of $L^2(\Omega)$ functions

This paragraph gives a review of the methods of construction of $S_\ell-$approximations of $L^2(\Omega)$ functions that will be used in this and the later chapters.

Definition 4.1.2. *(L^2-projection)*
For any $v \in L^2(\Omega)$, the L^2-projection onto S_ℓ, $P_{\ell,L^2} : L^2(\Omega) \to S_\ell$, is defined by

$$(P_{\ell,L^2}v, w)_{0,\Omega} = (v, w)_{0,\Omega}, \quad \forall\ w \in S_\ell$$

The Clément quasi-interpolation operator uses local L^2-projections to approximate the nodal values of functions.

Definition 4.1.3. *(Clément quasi-interpolation operator)*
The quasi-interpolation operator of Clément, $P_{\ell,C} : L^2(\Omega) \to S_\ell$, is defined as follows. Let $v \in L^2(\Omega)$ be given. For any $a \in \overline{\mathcal{N}_\ell}$, let $\pi_{\ell,a}$ be the $L^2(\omega_\ell^a)$-projection of v onto $P_1(\omega_\ell^a)$, i.e.,

$$\pi_{\ell,a}v \in P_1(\omega_\ell^a), \quad (v, w)_{0,\omega_\ell^a} = (\pi_{\ell,a}v, w)_{0,\omega_\ell^a}, \quad \forall\, w \in P_1(\omega_\ell^a).$$

The operator $P_{\ell,C}$ is uniquely defined by the relations

$$P_{\ell,C}v(a) = \pi_{\ell,a}v(a), \quad a \in \overline{\mathcal{N}}_\ell.$$

Remark 4.1.4. If $v \in V$, the above definition does not preserve the homogeneous boundary conditions. In order to obtain $P_{\ell,C}v \in V_\ell$, one could set

$$P_{\ell,C}v(a) = \begin{cases} \pi_{\ell,a}v(a), & a \in \mathcal{N}_\ell \\ 0 \qquad\quad , & a \in \mathcal{N}_\ell^\Gamma \end{cases}.$$

(see [61], p. 10).

Definition 4.1.5. *(Scott-Zhang interpolation operator)*
For each nodal point $a \in N_\ell$, let T stand for an arbitrary but fixed element from the patch ω_ℓ^a. Restrict the nodal basis function $\varphi_\ell^{(a)}$ to T according to

$$\varphi_T^{(a)} := \varphi_\ell^{(a)}|_T$$

and by $\{\phi_T^{(a)},\ a \in \mathcal{N}_\ell(T)\}$ denote the $L^2(T)$-dual basis of $\{\varphi_T^{(a)},\ a \in \mathcal{N}_\ell(T)\}$, i.e.,

$$\int_T \varphi_T^{(a)}\phi_T^{(p)}dx = \delta_{ap},\ \forall\ a, p \in \mathcal{N}_\ell(T).$$

The Scott-Zhang interpolation operator $P_{\ell,SZ} : L^2(\Omega) \to V_\ell$ is defined as

$$P_{\ell,SZ}v := \sum_{a\in\mathcal{N}_\ell} P_{\ell,SZ}v(a)\varphi_\ell^{(a)},$$

with the nodal coefficients

$$P_{\ell,SZ}v(a) := \int_T \phi_T^{(a)}(x)v(x)dx.$$

The Clément quasi-interpolation and the Scott-Zhang interpolation operators will be used in the a posteriori error analysis, whereas the L^2-projection will be employed for the discretization of the data functions of the optimal control problem. Let us recall some relevant properties of the above defined operators.

Proposition 4.1.6. *Let $P_\ell \in \{P_{\ell,C}, P_{\ell,SZ}, P_{\ell,L^2}\}$. There holds:*

(i) for all $0 \leq k \leq 2$, $0 \leq m \leq k$, and $v \in H^k(T)$ there holds $\|v - P_\ell v\|_{m,T} \lesssim h_T^{k-m}|v|_{k,T}$;

(ii) if $h_\ell \to 0$ for $\ell \to \infty$, then $\|v - P_\ell v\|_{0,\Omega} \to 0$ for $\ell \to \infty$ and $\forall\, v \in L^2(\Omega)$.

Proof. (i): The reader is referred to [19], p. 79, [57], p. 490, and to [15], p. 467. (ii): The proof essentially relies on the statement (i) (see [19], p. 82). □

Proposition 4.1.7. *(Some properties of $P_{\ell,SZ}$)*

(i) the Scott-Zhang operator is an interpolation operator, i.e., $P_{\ell,SZ} v_\ell = v_\ell$, $\forall v_\ell \in S_\ell$;

(ii) there holds $|P_{\ell,SZ} v(a)| \lesssim h_T^{-1}\|v\|_{0,T}$, $a \in \mathcal{N}_\ell(T)$, $T \in \mathcal{T}_\ell$, $v \in L^2(\Omega)$.

Proof. For the proof of *(i)* the reader is referred to [57], p. 487. Here only the proof of *(ii)* is given. Due to [57], p. 488, for any $a \in \mathcal{N}_\ell(T)$ and the associated element $T \in \mathcal{T}_\ell$, we have $\|\phi_T^{(a)}\|_{L^\infty(T)} \lesssim h_T^{-2}$. Thus, by the definition of $P_{\ell,SZ}$, for any pair $a \in \mathcal{N}_\ell(T)$ and $T \in \mathcal{T}_\ell$,

$$|P_{\ell,SZ} v(a)| = |\int_T \phi_T^{(a)}(x) v(x) dx| \leq \|\phi_T^{(a)}\|_{L^\infty(T)} \|v\|_{L^1(T)} \lesssim h_T^{-2}\|v\|_{L^1(T)} \lesssim h_T^{-1}\|v\|_{0,T}.$$

□

Proposition 4.1.8. *(Some properties of $P_{\ell,C}$)*
Let $v \in V$, $T \in \mathcal{T}_\ell$, and $E \in \mathcal{E}_\ell$ be arbitrary. The operator $P_{\ell,C}$ satisfies the following local error estimates.

$$\|v - P_{\ell,C} v\|_{0,T} \lesssim h_T \|v\|_{1,\omega_\ell^T}, \tag{4.1.3a}$$

$$\|v - P_{\ell,C} v\|_{0,E} \lesssim h_T^{1/2} \|v\|_{1,\omega_\ell^E}. \tag{4.1.3b}$$

Proof. The reader is referred to [61], [19], and [18], p. 145. □

It is emphasized that $P_{\ell,C}$ is not an interpolation operator in the sense of Proposition 4.1.7, (i).

4.2 Discretization of the problem

For the finite element approximation of the optimal control problem (3.1.1), the obstacle ψ and the shift control u^d are substituted by their interpolants $\psi_\ell \in V_\ell$ and $u_\ell^d \in S_\ell$, respectively (any of the nodal approximations described in the previous section can be employed in this case). To obtain nodal approximations for the functions f and y^d, we deliberately choose the method of L^2-projection, forcing the interpolants $y_\ell^d, f_\ell \in S_\ell$ to conform to the conditions

$$(f_\ell, v_\ell)_{0,\Omega} = (f, v_\ell)_{0,\Omega}, \quad (y_\ell^d, v_\ell)_{0,\Omega} = (y^d, v_\ell)_{0,\Omega}, \quad \forall\, v_\ell \in V_\ell. \tag{4.2.1}$$

Approximating the state $y \in V$ and the control $u \in L^2(\Omega)$ by finite element functions $y_\ell \in V_\ell$ and $u_\ell \in S_\ell$, the discrete optimal control problem is given as follows:

$$\begin{aligned}
\textit{minimize} \quad & J_\ell(y_\ell, u_\ell) \;:=\; \frac{1}{2}\,\|y_\ell - y_\ell^d\|_{0,\Omega}^2 \;+\; \frac{\alpha}{2}\,\|u_\ell - u_\ell^d\|_{0,\Omega}^2 && \text{(4.2.2a)}\\
\textit{over} \quad & (y_\ell, u_\ell) \in V_\ell \times S_\ell, \\
\textit{subject to} \quad & a(y_\ell, y_\ell - v_\ell) \;\le\; (f_\ell + u_\ell, y_\ell - v_\ell)_{0,\Omega}\,,\ \forall\, v_\ell \in K_\ell, && \text{(4.2.2b)}\\
& y_\ell \in K_\ell = \{v_\ell \in V_\ell \mid v_\ell(x) \le \psi_\ell(x),\ x \in \Omega\}.
\end{aligned}$$

We refer to J_ℓ and K_ℓ as the discrete objective functional and the discrete constraint set and to y_ℓ and u_ℓ as the discrete state and the discrete control.

Denoting by $\mathcal{S}_\ell : S_\ell \to V_\ell$ the discrete control-to-state mapping which assigns to a control $u_\ell \in S_\ell$ the unique solution $y_\ell \in V_\ell$ of the discrete variational inequality (4.2.2b), the control-reduced form of (4.2.2) reads:

$$\begin{aligned}
\textit{minimize} \quad & J_\ell^{red}(u_\ell) \;:=\; \frac{1}{2}\,\|\mathcal{S}_\ell u_\ell - y_\ell^d\|_{0,\Omega}^2 \;+\; \frac{\alpha}{2}\,\|u_\ell - u_\ell^d\|_{0,\Omega}^2 && \text{(4.2.3)}\\
\textit{over} \quad & u_\ell \in S_\ell.
\end{aligned}$$

Theorem 4.2.1. *The discrete optimal control problem* (4.2.2) *admits an optimal solution* $(y_\ell, u_\ell) \in V_\ell \times S_\ell$*. In general, there is no uniqueness.*

Proof. The proof follows the same lines as that of Theorem 3.1.1. □

As in the continuous case, by introducing a slack variable $\sigma_\ell \in \mathcal{M}_\ell$, the discrete optimal control problem (4.2.2) can be equivalently reformulated as the discrete complementarity problem:

$$\begin{aligned}
\textit{minimize} \quad & J_\ell(y_\ell, u_\ell) \;=\; \frac{1}{2}\,\|y_\ell - y_\ell^d\|_{0,\Omega}^2 \;+\; \frac{\alpha}{2}\,\|u_\ell - u_\ell^d\|_{0,\Omega}^2 && \text{(4.2.4a)}\\
\textit{over} \quad & (y_\ell, u_\ell) \in V_\ell \times S_\ell, \\
\textit{subject to} \quad & a(y_\ell, v_\ell) = (f_\ell + u_\ell, v_\ell)_{0,\Omega} - \langle\langle \sigma_\ell, v_\ell \rangle\rangle,\ \forall\, v_\ell \in V_\ell, && \text{(4.2.4b)}\\
& y_\ell \in K_\ell,\ \sigma_\ell \in \mathcal{M}_\ell \cap \mathcal{M}_+(\overline{\Omega})\,,\ \langle\langle \sigma_\ell, y_\ell - \psi_\ell \rangle\rangle = 0. && \text{(4.2.4c)}
\end{aligned}$$

4.3 Localization of the discrete complementarity constraint

As in the continuous setting, based on the conditions $y_\ell \in K_\ell$ and $\sigma_\ell \in \mathcal{M}_\ell \cap \mathcal{M}_+(\overline{\Omega})$, two partitions of $\overline{\Omega}$ are introduced. In either case, the computational domain is split into two subsets - the subset where the determining inequality holds strictly and the subset where it holds as an equality. The first condition gives rise to the discrete active and the inactive set, the second one - to the discrete strongly active and zero set.

4.3.1 Discrete active and inactive set

Definition 4.3.1. *(Zero and non-zero nodal points w.r.t. $v_\ell \in V_\ell$)*
Any $v_\ell \in V_\ell$ yields a partition

$$\overline{\mathcal{N}}_\ell = z(v_\ell) \cup c(v_\ell),$$

where $z(v_\ell) := \{a \in \overline{\mathcal{N}}_\ell \mid v_\ell(a) = 0\}$ is the set of zero nodal points with respect to v_ℓ and $c(v_\ell) := \overline{\mathcal{N}}_\ell \setminus z(v_\ell)$ is the set of non-zero nodal points with respect to v_ℓ.

Further, we split the elements of $\mathcal{T}_\ell$ into the sets of purely zero, non-zero, and mixed triangles with respect to $v_\ell \in V_\ell$.

Definition 4.3.2. *(Zero, non-zero, and mixed triangles w.r.t. $v_\ell \in V_\ell$)*
For any $v_\ell \in V_\ell$, we partition $\mathcal{T}_\ell$ into the sets of zero, non-zero, and mixed triangles with respect to v_ℓ,

$$\mathcal{T}_\ell = \mathcal{T}_\ell^z(v_\ell) \cup \mathcal{T}_\ell^c(v_\ell) \cup \mathcal{T}_\ell^m(v_\ell),$$

where

$$\begin{aligned}
\mathcal{T}_\ell^z(v_\ell) &:= \{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subseteq z(v_\ell)\},\\
\mathcal{T}_\ell^c(v_\ell) &:= \{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subseteq c(v_\ell)\},\\
\mathcal{T}_\ell^m(v_\ell) &:= \mathcal{T}_\ell \setminus (\mathcal{T}_\ell^z(v_\ell) \cup \mathcal{T}_\ell^c(v_\ell))\,.
\end{aligned}$$

Using the latter notation, we partition $\mathcal{N}_\ell$ into the sets of active and inactive nodal points, $\mathcal{E}_\ell$ into the sets of active, inactive, and free boundary edges with respect to y_ℓ. Moreover, every triangle in $\mathcal{T}_\ell$ is marked as active, inactive, or free boundary with respect to y_ℓ, contributing to the discrete active set, discrete inactive set, and discrete free boundary with respect to y_ℓ, respectively.

Definition 4.3.3. *(Active and inactive nodal points)*
We refer to $A_\ell := z(\psi_\ell - y_\ell) \cap \mathcal{N}_\ell$ and $I_\ell := c(\psi_\ell - y_\ell) \cap \mathcal{N}_\ell$ as the sets of active and inactive nodal points, respectively.

Definition 4.3.4. *(Active and inactive edges, free boundary edges w.r.t. y_ℓ)*
We partition the set $\mathcal{E}_\ell$ into mutually non-intersecting subsets

$$\mathcal{E}_\ell = \mathcal{E}_{\mathcal{A}_\ell} \cup \mathcal{E}_{\mathcal{F}_\ell(y_\ell)} \cup \overset{\circ}{\mathcal{E}}_{\mathcal{I}_\ell},$$

where $\mathcal{E}_{\mathcal{A}_\ell} := \{E \in \mathcal{E}_\ell \,|\, \mathcal{N}_\ell(E) \subseteq A_\ell\}$, $\overset{\circ}{\mathcal{E}}_{\mathcal{I}_\ell} := \{E \in \mathcal{E}_\ell \,|\, \mathcal{N}_\ell(E) \subseteq I_\ell\}$, and $\mathcal{E}_{\mathcal{F}_\ell(y_\ell)} := \mathcal{E}_\ell \setminus (\mathcal{E}_{\mathcal{A}_\ell} \cup \overset{\circ}{\mathcal{E}}_{\mathcal{I}_\ell})$ are the sets of active edges, purely inactive edges, and free boundary edges with respect to y_ℓ, respectively. The set of inactive edges is defined by $\mathcal{E}_{\mathcal{I}_\ell} := \overset{\circ}{\mathcal{E}}_{\mathcal{I}_\ell} \cup \mathcal{E}_{\mathcal{F}_\ell(y_\ell)}$.

Definition 4.3.5. *(Discrete active set, discrete inactive set, discrete free boundary w.r.t. y_ℓ)*
The constraint $y_\ell \in K_\ell$ yields a partition

$$\overline{\Omega} = \mathcal{A}_\ell \cup \mathcal{F}_\ell(y_\ell) \cup \overset{\circ}{\mathcal{I}}_\ell,$$

where

(i) $\mathcal{A}_\ell := \bigcup\{T \in \mathcal{T}_\ell^z(\psi_\ell - y_\ell)\}$ *is the discrete active set,*

(ii) $\overset{\circ}{\mathcal{I}}_\ell := \bigcup\{T \in \mathcal{T}_\ell^c(\psi_\ell - y_\ell)\}$ *is the discrete purely inactive set,*

(iii) $\mathcal{F}_\ell(y_\ell) := \bigcup\{T \in \mathcal{T}_\ell^m(\psi_\ell - y_\ell)\}$ *is the discrete free boundary with respect to y_ℓ.*

The discrete inactive set is defined by $\mathcal{I}_\ell := \overset{\circ}{\mathcal{I}}_\ell \cup \mathcal{F}_\ell(y_\ell)$. A triangle T is called active if $T \in \mathcal{T}_\ell(\mathcal{A}_\ell)$, inactive if $T \in \mathcal{T}_\ell(\mathcal{I}_\ell)$ and free boundary with respect to y_ℓ if $T \in \mathcal{T}_\ell(\mathcal{F}_\ell(y_\ell))$.

Definition 4.3.6. *(Isolated active nodal points and edges)*
An active nodal point $a \in A_\ell$ is said to be isolated if $\mathcal{N}_\ell(\omega_\ell^a) \setminus \{a\} \subseteq I_\ell \cup \mathcal{N}_\ell^\Gamma$. An active edge $E \in \mathcal{E}_{\mathcal{A}_\ell}$ is called isolated if $E \in \mathcal{E}_{\mathcal{A}_\ell} \setminus \mathcal{E}_\ell(\mathcal{A}_\ell)$.

In other words, a nodal point $a \in A_\ell$ is isolated if there is no $E \in \mathcal{E}_{\mathcal{A}_\ell}$ such that $a \in \mathcal{N}_\ell(E)$, whereas an active edge $E \in \mathcal{E}_{\mathcal{A}_\ell}$ is isolated if there is no $T \in \mathcal{T}_\ell(\mathcal{A}_\ell)$ such that $E \in \mathcal{E}_\ell(T)$. Obviously, all isolated active nodal points and edges lie within the set $\mathcal{F}_\ell(y_\ell)$.

4.3.2 Discrete strongly active and zero sets

The elements of $\mathcal{M}_\ell$ are defined *only* in the interior nodal points. This fact is taken into account in the following system of definitions.

Definition 4.3.7. *(Zero and non-zero nodal points w.r.t. $\lambda_\ell \in \mathcal{M}_\ell$)*
Any $\lambda_\ell \in \mathcal{M}_\ell$ yields a partition

$$\mathcal{N}_\ell = z(\lambda_\ell) \cup c(\lambda_\ell),$$

where $z(\lambda_\ell) := \{a \in \mathcal{N}_\ell \mid \lambda_\ell(a) = 0\}$ is the set of zero nodal points with respect to λ_ℓ and $c(\lambda_\ell) := \mathcal{N}_\ell \setminus z(\lambda_\ell)$ is the set of non-zero nodal points with respect to λ_ℓ.

As $\lambda_\ell \in \mathcal{M}_\ell$ has no values in $\mathcal{N}_\ell^\Gamma$ (cf. Remark 4.1.1), the boundary nodes are not included into the classification introduced in the latter definition. A classification of triangles based on the nodal values of $\lambda_\ell \in \mathcal{M}_\ell$ comparable to the one introduced in the Definition 4.3.2, would require special agreements for the triangles adjacent to the boundary. In the following definition, we take a decision upon the type of T, $T \in \mathcal{T}_\ell \setminus \mathcal{T}_\ell(\Omega)$ depending on whether the condition $T \subseteq \mathcal{A}_\ell$ is satisfied. As we shall see later on, this condition helps to transfer some distinctive features of the continuous sets into the discrete setting (cf. Proposition 2.4.9 in the continuous case and Proposition 4.3.14 in the discrete case).

Definition 4.3.8. *(Zero, non-zero, and mixed triangles w.r.t. $\lambda_\ell \in \mathcal{M}_\ell$)*
For any $\lambda_\ell \in \mathcal{M}_\ell$, we partition $\mathcal{T}_\ell$ into the sets of zero, non-zero, and mixed triangles with respect to λ_ℓ,

$$\mathcal{T}_\ell = \mathcal{T}_\ell^z(\lambda_\ell) \cup \mathcal{T}_\ell^c(\lambda_\ell) \cup \mathcal{T}_\ell^m(\lambda_\ell),$$

where

$$\begin{aligned}
\mathcal{T}_\ell^z(\lambda_\ell) &:= \{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subseteq z(\lambda_\ell) \cup \mathcal{N}_\ell^\Gamma\},\\
\mathcal{T}_\ell^c(\lambda_\ell) &:= \{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subseteq c(\lambda_\ell)\} \cup\\
&\qquad \cup \{T \in \mathcal{T}_\ell \setminus \mathcal{T}_\ell(\Omega) \mid \varnothing \neq \mathcal{N}_\ell(T) \cap \mathcal{N}_\ell \subseteq c(\lambda_\ell) \wedge T \subseteq \mathcal{A}_\ell\},\\
\mathcal{T}_\ell^m(\lambda_\ell) &:= \mathcal{T}_\ell \setminus (\mathcal{T}_\ell^z(\lambda_\ell) \cup \mathcal{T}_\ell^c(\lambda_\ell))\,.
\end{aligned}$$

We further define the sets of strongly active and zero nodal points, the sets of strongly active, zero and free boundary edges and triangles, as well as the discrete strongly active set, the discrete zero set, and the discrete free boundary with respect to σ_ℓ.

Definition 4.3.9. *(Strongly active and zero nodal points)*
We refer to $Z_\ell := z(\sigma_\ell)$ and $C_\ell := c(\sigma_\ell)$ as the sets of zero and and strongly active nodal points. A strongly active nodal point $a \in C_\ell$ is said to be isolated if $\mathcal{N}_\ell(\omega_\ell^a) \setminus \{a\} \subseteq Z_\ell \cup \mathcal{N}_\ell^\Gamma$.

Definition 4.3.10. *(Strongly active and zero edges, free boundary edges w.r.t. σ_ℓ)*
We partition the set $\mathcal{E}_\ell$ into mutually non-intersecting subsets

$$\mathcal{E}_\ell = \mathcal{E}_{\mathcal{C}_\ell} \cup \mathcal{E}_{\mathcal{F}_\ell(\sigma_\ell)} \cup \overset{\circ}{\mathcal{E}}_{\mathcal{Z}_\ell},$$

where $\mathcal{E}_{\mathcal{C}_\ell} := \{E \in \mathcal{E}_\ell \,|\, \mathcal{N}_\ell(E) \subseteq C_\ell\}$, $\overset{\circ}{\mathcal{E}}_{\mathcal{Z}_\ell} := \{E \in \mathcal{E}_\ell \,|\, \mathcal{N}_\ell(E) \subseteq Z_\ell\}$, and $\mathcal{E}_{\mathcal{F}_\ell(\sigma_\ell)} := \mathcal{E}_\ell \setminus (\mathcal{E}_{\mathcal{C}_\ell} \cup \overset{\circ}{\mathcal{E}}_{\mathcal{Z}_\ell})$ are the sets of strongly active edges, purely zero edges, and free boundary edges with respect to σ_ℓ, respectively. The set of zero edges is defined by $\mathcal{E}_{\mathcal{Z}_\ell} := \overset{\circ}{\mathcal{E}}_{\mathcal{Z}_\ell} \cup \mathcal{E}_{\mathcal{F}_\ell(\sigma_\ell)}$.

Definition 4.3.11. *(Discrete strongly active set, discrete zero set, discrete free boundary w.r.t. σ_ℓ)*
The constraint $\sigma_\ell \in \mathcal{M}_\ell \cap \mathcal{M}_+(\Omega)$ yields a partition

$$\overline{\Omega} = \mathcal{C}_\ell \cup \mathcal{F}_\ell(\sigma_\ell) \cup \overset{\circ}{\mathcal{Z}}_\ell,$$

where

(*i*) $\mathcal{C}_\ell := \bigcup\{T \in \mathcal{T}_\ell^c(\sigma_\ell)\}$ *is the discrete strongly active set,*

(*ii*) $\overset{\circ}{\mathcal{Z}}_\ell := \bigcup\{T \in \mathcal{T}_\ell^z(\sigma_\ell)\}$ *is the discrete purely zero set,*

(*iii*) $\mathcal{F}_\ell(\sigma_\ell) := \bigcup\{T \in \mathcal{T}_\ell^{zc}(\sigma_\ell)\}$ *is the discrete free boundary with respect to σ_ℓ.*

The discrete zero set is defined by $\mathcal{Z}_\ell := \overset{\circ}{\mathcal{Z}}_\ell \cup \mathcal{F}_\ell(\sigma_\ell)$. A triangle T is called strongly active if $T \in \mathcal{T}_\ell(\mathcal{C}_\ell)$, inactive if $T \in \mathcal{T}_\ell(\mathcal{Z}_\ell)$ and free boundary with respect to y_ℓ if $T \in \mathcal{T}_\ell(\mathcal{F}_\ell(\sigma_\ell))$.

Definition 4.3.12. *(Isolated strongly active nodal points and edges)*
A strongly active nodal point $a \in C_\ell$ is said to be isolated if $\mathcal{N}_\ell(\omega_\ell^a) \setminus \{a\} \subseteq Z_\ell \cup \mathcal{N}_\ell^\Gamma$. A strongly active edge $E \in \mathcal{E}_{\mathcal{C}_\ell}$ is called isolated if $E \in \mathcal{E}_{\mathcal{C}_\ell} \setminus \mathcal{E}_\ell(\mathcal{C}_\ell)$.

All the isolated strongly active nodal points and edges are contained within $\mathcal{F}_\ell(\sigma_\ell)$.

4.3.3 Discrete biactive set and the concept of strict complementarity

In this paragraph, we extend the concepts of *biactive set*, *strict complementarity* and *lack of strict complementarity* to the discrete setting.

Definition 4.3.13. *(Biactive nodal points, discrete biactive set, strict complementarity, lack of strict complementarity)*

(*i*) $B_\ell := A_\ell \cap Z_\ell$ *is the set of biactive nodal points;*

(*ii*) $\mathcal{B}_\ell := cl(\mathcal{A}_\ell \setminus \mathcal{C}_\ell)$ *is the discrete biactive set;*

(*iii*) *If $B_\ell = \varnothing$, it is said that the strict complementarity condition is satisfied at the discrete solution;*

(*iv*) *if $B_\ell \neq \varnothing$, it is said that there is lack of strict complementarity at the discrete solution.*

4.3.4 Properties of the discrete sets

Proposition 4.3.14. *The discrete sets defined in the previous paragraphs satisfy:*

(*i*) $C_\ell \subseteq A_\ell$ *and* $\mathcal{C}_\ell \subseteq \mathcal{A}_\ell$*;*

(*ii*) *If* $B_\ell = \varnothing$ *and* $\{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subseteq \mathcal{N}_\ell^\Gamma\} = \varnothing$*, then* $\mathcal{B}_\ell = \varnothing$ *(i.e.,* $\mathcal{A}_\ell = \mathcal{C}_\ell$*).*

Proof. (*i*): Due to the complementarity between $\psi_\ell - y_\ell$ and σ_ℓ, there holds $C_\ell \subseteq A_\ell$. The second statement is evident for the interior triangles $\mathcal{T}_\ell(\Omega)$, we only need to consider the triangles in $\mathcal{T}_\ell \backslash \mathcal{T}_\ell(\Omega)$. By the definition of $\mathcal{C}_\ell$ and due to the complementarity $\psi_\ell - y_\ell$ and σ_ℓ, if $T \in \mathcal{T}_\ell \setminus \mathcal{T}_\ell(\Omega)$ and $T \subseteq \mathcal{C}_\ell$, there holds $T \subseteq \mathcal{A}_\ell$.
(*ii*) Similar to (*i*). □

We might encounter meshes with $\{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subseteq \mathcal{N}_\ell^\Gamma\} \neq \varnothing$. In such a case, the assumptions of Proposition 4.3.14 (*ii*) can be fulfilled on the following mesh levels by refining the elements of the troublesome set.

Remark 4.3.15.

(*i*) One can show that

$$\bigcup_{E \in \mathcal{E}_{\mathcal{Z}_\ell}} \omega_E = \mathcal{Z}_\ell, \qquad \bigcup_{E \in \mathcal{E}_{\mathcal{I}_\ell}} \omega_E = \mathcal{I}_\ell. \tag{4.3.1}$$

(*ii*) When σ_ℓ is considered as a functional on V_ℓ, there holds

$$\mathcal{O}_{\sigma_\ell} = \operatorname{int}(\mathcal{Z}_\ell), \quad \operatorname{supp}(\sigma_\ell) = \mathcal{C}_\ell,$$

(see Definition 2.3.5) whereas if understood as an element of $\mathcal{M}(\overline{\Omega})$,

$$\mathcal{O}_{\sigma_\ell} = Z_\ell, \quad \operatorname{supp}(\sigma_\ell) = C_\ell.$$

(*iii*) In the continuous setting, the sets $\mathcal{A}$, $\mathcal{I}$, $\mathcal{C}$, and $\mathcal{Z}$ are open and the free boundary is a 1-dimensional manifold. Discrete sets of the same quality can be constructed as follows:

$$\begin{aligned} \mathcal{F}_\ell^*(\sigma_\ell) &:= (\mathcal{F}_\ell(\sigma_\ell) \cap \mathcal{C}_\ell) \cap \Omega, & \mathcal{C}_\ell^* &:= \mathcal{C}_\ell \cap \Omega, & \mathcal{Z}_\ell^* &:= \mathcal{Z}_\ell \cap \Omega, \\ \mathcal{F}_\ell^*(y_\ell) &:= (\mathcal{F}_\ell(y_\ell) \cap \mathcal{A}_\ell) \cap \Omega, & \mathcal{A}_\ell^* &:= \mathcal{A}_\ell \cap \Omega, & \mathcal{I}_\ell^* &:= \mathcal{I}_\ell \cap \Omega. \end{aligned}$$

These sets, too, satisfy $\mathcal{C}_\ell^* \subseteq \mathcal{A}_\ell^*$, and $\mathcal{C}_\ell^* = \mathcal{A}_\ell^*$, as soon as $B_\ell = \varnothing$ and $\{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subset \mathcal{N}_\ell^\Gamma\} = \varnothing$.

4.4 First order optimality conditions

From the theory of mathematical problems with complementarity constraints (MPCC) in finite dimensional spaces, it follows that our discrete problem always has a strongly stationary point. Here, a brief review of the results of [56] is given, that supports this fact.

4.4.1 MPCC in finite dimensions: constraint qualifications and stationarity concepts

Let us consider an MPCC in the form

$$\begin{aligned} \textit{minimize} \quad & f(\boldsymbol{z}) \\ \textit{over} \quad & \boldsymbol{z} \in \mathbb{R}^s, \\ \textit{subject to} \quad & \min\{(\boldsymbol{F}_1(\boldsymbol{z}))_k, (\boldsymbol{F}_2(\boldsymbol{z}))_k\} = 0, \quad k = 1, \dots, m, && \text{(4.4.1a)} \\ & \boldsymbol{g}(\boldsymbol{z}) \leq 0, && \text{(4.4.1b)} \\ & \boldsymbol{h}(\boldsymbol{z}) = 0, && \text{(4.4.1c)} \end{aligned}$$

where $f : \mathbb{R}^s \to \mathbb{R}$, $\boldsymbol{g} : \mathbb{R}^s \to \mathbb{R}^p$, $\boldsymbol{h} : \mathbb{R}^s \to \mathbb{R}^q$, and $\boldsymbol{F}_k : \mathbb{R}^s \to \mathbb{R}^m$, $k = 1, 2$ are assumed to be smooth and there holds $m \geq 1$. The non-smooth constraint (4.4.1a) is an equivalent form of the complementarity constraint

$$\boldsymbol{F}_1(\boldsymbol{z}) \geq 0, \quad \boldsymbol{F}_2(\boldsymbol{z}) \geq 0, \quad \boldsymbol{F}_1(\boldsymbol{z}) \cdot \boldsymbol{F}_2(\boldsymbol{z}) = 0.$$

Note that (4.4.1a) allows direct application of classical constraint qualification concepts, unlike the constraint above (see [45], p. 14). The Lagrangian functional associated with problem (4.4.1) is given by

$$L(\boldsymbol{z}, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\lambda}_g, \boldsymbol{\lambda}_h) = f(\boldsymbol{z}) - \boldsymbol{F}_1(\boldsymbol{z}) \cdot \boldsymbol{\Gamma}_1 - \boldsymbol{F}_2(\boldsymbol{z}) \cdot \boldsymbol{\Gamma}_2 + \boldsymbol{g}(\boldsymbol{z}) \cdot \boldsymbol{\lambda}_g + \boldsymbol{h}(\boldsymbol{z}) \cdot \boldsymbol{\lambda}_h,$$

where $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2 \in \mathbb{R}^m$, $\boldsymbol{\lambda}_g \in \mathbb{R}^p$, and $\boldsymbol{\lambda}_h \in \mathbb{R}^q$ are the associated Lagrangian multipliers. Let us recall two types of constraint qualifications for MPCC that are employed in [56].

Definition 4.4.1. *(MFCQ and SMFCQ)*

(*i*) *A feasible point* $\boldsymbol{z}$ *of* (4.4.1) *satisfies the Mangasarian-Fromovitz constraint qualification (MFCQ) if*

(i.1) *the gradients*

$$\begin{aligned} &\nabla(\boldsymbol{F}_k(\boldsymbol{z}))_i, \ \textit{for } (i,k) \in \{1, \dots, m\} \times \{1, 2\} \textit{ with } (\boldsymbol{F}_k(\boldsymbol{z}))_i = 0, \\ &\nabla(\boldsymbol{h}(\boldsymbol{z}))_i, \ i = 1, \dots, q. \end{aligned}$$

are linearly independent;

(i.2) *there exists a vector* $\boldsymbol{v} \in \mathbb{R}^s$ *orthogonal to these gradients, such that*

$$\nabla(\boldsymbol{g}(\boldsymbol{z}))_i \cdot \boldsymbol{v} < 0, \ \textit{for } i \in \{1, \dots, p\} \textit{ with } (\boldsymbol{g}(\boldsymbol{z}))_i = 0.$$

(*ii*) *A feasible point* $\boldsymbol{z} \in \mathbb{R}^s$ *of* (4.4.1) *satisfies the strong Mangasarian-Fromovitz constraint qualification (SMFCQ) if there exist Lagrange multipliers* $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, $\boldsymbol{\lambda}_g$, *and* $\boldsymbol{\lambda}_h$ *such that*

(ii.1) *the gradients*

$$\begin{aligned}&\nabla(\boldsymbol{F}_k(\boldsymbol{z}))_i, \text{ for } (i,k) \in \{1,\dots,m\} \times \{1,2\} \text{ with } (\boldsymbol{F}_k(\boldsymbol{z}))_i = 0,\\ &\nabla(\boldsymbol{h}(\boldsymbol{z}))_i, \ i = 1,\dots,q,\\ &\nabla(\boldsymbol{g}(\boldsymbol{z}))_i \text{ for } i \in \{1,\dots,p\} \text{ with } (\boldsymbol{\lambda}_g)_i > 0.\end{aligned}$$

are linearly independent;

(ii.2) *there exists a vector* $\boldsymbol{v} \in \mathbb{R}^s$ *orthogonal to these gradients, such that*

$$\nabla(\boldsymbol{g}(\boldsymbol{z}))_i \cdot \boldsymbol{v} < 0, \text{ for } i \in \{1,\dots,p\} \text{ with } (\boldsymbol{g}(\boldsymbol{z}))_i = (\boldsymbol{\lambda}_g)_i = 0.$$

In connection with these constraint classifications, two kinds of systems of necessary optimality conditions for problem (4.4.1) can be introduced.

Theorem 4.4.2. *Let* $\boldsymbol{z} \in \mathbb{R}^s$ *be a local minimizer of* (4.4.1).

(i) *If MFCQ holds at* $\boldsymbol{z}$*, then there exist multipliers* $\boldsymbol{\Gamma}_1$*,* $\boldsymbol{\Gamma}_2$*,* $\boldsymbol{\lambda}_g$*, and* $\boldsymbol{\lambda}_h$ *such that*

$$\nabla_{\boldsymbol{z}} L(\boldsymbol{z}, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\lambda}_g, \boldsymbol{\lambda}_h) = \boldsymbol{0}_s, \tag{4.4.2a}$$
$$\boldsymbol{F}_k(\boldsymbol{z}) * \boldsymbol{\Gamma}_k = \boldsymbol{0}_m, \ k = 1,2, \tag{4.4.2b}$$
$$(\boldsymbol{\Gamma}_1)_i(\boldsymbol{\Gamma}_2)_i \geq 0 \text{ for } i \in \{1,\dots,m\} \text{ with } (\boldsymbol{F}_1(\boldsymbol{z}))_i = (\boldsymbol{F}_2(\boldsymbol{z}))_i = 0, \tag{4.4.2c}$$
$$\boldsymbol{\lambda}_g \geq \boldsymbol{0}_p, \ \boldsymbol{g}(\boldsymbol{z}) \cdot \boldsymbol{\lambda}_g = 0. \tag{4.4.2d}$$

(ii) *If SMFCQ holds at* $\boldsymbol{z}$*, then there exist unique multipliers* $\boldsymbol{\Gamma}_1$*,* $\boldsymbol{\Gamma}_2$*,* $\boldsymbol{\lambda}_g$*, and* $\boldsymbol{\lambda}_h$ *such that*

$$\nabla_{\boldsymbol{z}} L(\boldsymbol{z}, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\lambda}_g, \boldsymbol{\lambda}_h) = \boldsymbol{0}_s, \tag{4.4.3a}$$
$$\boldsymbol{F}_k(\boldsymbol{z}) * \boldsymbol{\Gamma}_k = \boldsymbol{0}_m, \ k = 1,2, \tag{4.4.3b}$$
$$(\boldsymbol{\Gamma}_1)_i \geq 0, (\boldsymbol{\Gamma}_2)_i \geq 0 \text{ for } i \in \{1,\dots,m\} \text{ with } (\boldsymbol{F}_1(\boldsymbol{z}))_i = (\boldsymbol{F}_2(\boldsymbol{z}))_i = 0, \tag{4.4.3c}$$
$$\boldsymbol{\lambda}_g \geq \boldsymbol{0}_p, \ \boldsymbol{g}(\boldsymbol{z}) \cdot \boldsymbol{\lambda}_g = 0. \tag{4.4.3d}$$

Proof. See [56], p. 7. □

This result allows to define the following stationarity concepts for the MPCC under consideration.

Definition 4.4.3. *(C-stationary and S-stationary points of* (4.4.1)*)*

(i) *A feasible point* $\boldsymbol{z} \in \mathbb{R}^s$ *of* (4.4.1) *is called C-stationary if there exist multipliers* $\boldsymbol{\Gamma}_1$*,* $\boldsymbol{\Gamma}_2$*,* $\boldsymbol{\lambda}_g$*, and* $\boldsymbol{\lambda}_h$ *such that the system of conditions* (4.4.2) *is satisfied.*

(ii) *A feasible point* $\boldsymbol{z} \in \mathbb{R}^s$ *of* (4.4.1) *is called S-stationary if there exist multipliers* $\boldsymbol{\Gamma}_1$*,* $\boldsymbol{\Gamma}_2$*,* $\boldsymbol{\lambda}_g$*, and* $\boldsymbol{\lambda}_h$ *such that the system of conditions* (4.4.3) *is satisfied.*

Remark 4.4.4.

(*i*) There are further types of stationarity concepts available for problem (4.4.1). For more details, the reader is referred to [56] and references therein.

(*ii*) Note that the conditions for S-stationarity correspond to the Karush-Kuhn-Tucker conditions for the so-called relaxed nonlinear problem corresponding to (4.4.1) (see [56]).

4.4.2 Theory of MPCC in finite dimensions in application to problem (4.2.4)

We next establish a connection between problems (4.4.1) and (4.2.4). This requires a reformulation of (4.2.4) in vector form. Let us start by introducing some further notation.

Suppose the vertices of $\mathcal{T}_\ell$ are enumerated so that $\mathcal{N}_\ell = \{a_1, \dots, a_{N_\ell}\}$ and $\mathcal{N}_\ell^\Gamma = \{a_{N_\ell+1}, \dots, a_{\overline{N}_\ell}\}$. We refer to $\boldsymbol{n}_\ell := (1, \dots, N_\ell)^T$, $\boldsymbol{n}_\ell^\Gamma := (N_\ell + 1, \dots, \overline{N}_\ell)^T$, and $\overline{\boldsymbol{n}}_\ell := (1, \dots, \overline{N}_\ell)^T$ as the vectors of interior, boundary, and overall nodal indices, respectively. Further, setting

$$N_\ell^A := \mathrm{card}(A_\ell),\ N_\ell^I := \mathrm{card}(I_\ell),\ N_\ell^C := \mathrm{card}(C_\ell),\ N_\ell^Z := \mathrm{card}(Z_\ell),\ N_\ell^B := \mathrm{card}(B_\ell),$$

we define the vectors $\boldsymbol{n}_\ell^A \in \mathbb{R}^{N_\ell^A}$, $\boldsymbol{n}_\ell^I \in \mathbb{R}^{N_\ell^I}$, $\boldsymbol{n}_\ell^C \in \mathbb{R}^{N_\ell^C}$, $\boldsymbol{n}_\ell^Z \in \mathbb{R}^{N_\ell^Z}$, and $\boldsymbol{n}_\ell^B \in \mathbb{R}^{N_\ell^B}$ as the vectors of indices of active, inactive, strongly active, zero, and biactive vertices, respectively. These vectors are sub-vectors of $\boldsymbol{n}_\ell$. Let us also employ a simplified notation for the basis functions spanning the finite element space S_ℓ,

$$\varphi_i := \varphi_\ell^{(a_i)}, \quad i \in \{1, \dots, \overline{N}_\ell\}.$$

For an arbitrary vector $\boldsymbol{q} \in \mathbb{R}^n$ and an arbitrary but fixed vector of indices $\boldsymbol{n}_r \in \mathbb{R}^{n_r}$ that contains n^r selected components of the vector $(1, \dots, n)^T$, we refer to a $\boldsymbol{n}_r$-subvector of $\boldsymbol{q}$ as

$$\boldsymbol{q}(\boldsymbol{n}_r) \in \mathbb{R}^{n_r}, \quad (\boldsymbol{q}(\boldsymbol{n}_r))_i := (\boldsymbol{q})_{(\boldsymbol{n}_r)_i}, \quad 1 \le i \le n_r.$$

For an arbitrary matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ and arbitrary but fixed vectors of indices $\boldsymbol{n}_r \in \mathbb{R}^{n_r}$ and $\boldsymbol{n}_c \in \mathbb{R}^{n_c}$ with $n_r, n_c \le n$ that contain n_r and n_c selected elements of the vector $(1, \dots, n)^T$, respectively, we refer to $\boldsymbol{Q}(\boldsymbol{n}_r, \boldsymbol{n}_c) \in \mathbb{R}^{n_r \times n_c}$ as a $(\boldsymbol{n}_r, \boldsymbol{n}_c)$-submatrix of $\boldsymbol{Q}$, whose elements are specified by the relations

$$(\boldsymbol{Q}(\boldsymbol{n}_r, \boldsymbol{n}_c))_{i,j} := (\boldsymbol{Q})_{(\boldsymbol{n}_r)_i, (\boldsymbol{n}_r)_j},\ 1 \le i \le n_r,\ 1 \le j \le n_c.$$

Functions $v_\ell \in V_\ell$, $s_\ell \in S_\ell$, and $\lambda_\ell \in \mathcal{M}_\ell$ are uniquely defined by their nodal values $v_\ell(a_i)$, $i = 1, \dots, N_\ell$, $s_\ell(a_i)$, $i = 1, \dots, \overline{N}_\ell$, and $\lambda_\ell(a_i)$, $i = 1, \dots, N_\ell$, via the representations

$$v_\ell = \sum_{i=1}^{N_\ell} v_\ell(a_i)\varphi_i, \quad s_\ell = \sum_{i=1}^{\overline{N}_\ell} s_\ell(a_i)\varphi_i, \quad \lambda_\ell = \sum_{i=1}^{N_\ell} \lambda_\ell(a_i)\varphi_i,$$

respectively. Thus, to any $v_\ell \in V_\ell$, $s_\ell \in S_\ell$, and $\lambda_\ell \in \mathcal{M}_\ell$ we can assign vectors of nodal values $\boldsymbol{v}_\ell \in \mathbb{R}^{N_\ell}$, $\boldsymbol{s}_\ell \in \mathbb{R}^{\overline{N}_\ell}$, and $\boldsymbol{\lambda}_\ell \in \mathbb{R}^{N_\ell}$ with

$$\begin{aligned}(\boldsymbol{v}_\ell)_i &:= v_\ell(a_i), \quad i = 1, \dots, N_\ell,\\ (\boldsymbol{s}_\ell)_i &:= s_\ell(a_i), \quad i = 1, \dots, \overline{N}_\ell,\\ (\boldsymbol{\lambda}_\ell)_i &:= \lambda_\ell(a_i), \quad i = 1, \dots, N_\ell.\end{aligned}$$

Finally, by $\boldsymbol{K}_\ell$, $\boldsymbol{M}_\ell \in \mathbb{R}^{\overline{N}_\ell \times \overline{N}_\ell}$ we denote the stiffness and the mass matrices whose elements are given by

$$(\boldsymbol{K}_{nm})_{i,j} = (\nabla\varphi_i, \nabla\varphi_j)_{0,\Omega}, \quad (\boldsymbol{M}_{nm})_{i,j} = (\varphi_i, \varphi_j)_{0,\Omega}, \quad i, j = 1, \dots, \overline{N}_\ell.$$

In this paragraph, we will employ a special notation for selected sub-matrices of $\boldsymbol{K}_\ell$ and $\boldsymbol{M}_\ell$:

$$\begin{aligned}&\boldsymbol{K}_{N_\ell N_\ell} := \boldsymbol{K}_\ell(\boldsymbol{n}_\ell, \boldsymbol{n}_\ell),\\ &\boldsymbol{M}_{\overline{N}_\ell \overline{N}_\ell} := \boldsymbol{M}_\ell(\overline{\boldsymbol{n}}_\ell, \overline{\boldsymbol{n}}_\ell), \quad \boldsymbol{M}_{N_\ell N_\ell} := \boldsymbol{M}_\ell(\boldsymbol{n}_\ell, \boldsymbol{n}_\ell),\\ &\boldsymbol{M}_{\overline{N}_\ell N_\ell} := \boldsymbol{M}_\ell(\overline{\boldsymbol{n}}_\ell, \boldsymbol{n}_\ell), \quad \boldsymbol{M}_{N_\ell \overline{N}_\ell} := \boldsymbol{M}_\ell(\boldsymbol{n}_\ell, \overline{\boldsymbol{n}}_\ell).\end{aligned}$$

It is easy to verify that the discrete objective functional can be rewritten in the form

$$\begin{aligned}J_\ell(y_\ell, u_\ell) = J_\ell(\boldsymbol{y}_\ell, \boldsymbol{u}_\ell) := \frac{1}{2}\boldsymbol{y}_\ell^T \boldsymbol{M}_{N_\ell N_\ell}\boldsymbol{y}_\ell - \boldsymbol{y}_\ell^T \boldsymbol{M}_{N_\ell \overline{N}_\ell}\boldsymbol{y}_\ell^d + \frac{1}{2}(\boldsymbol{y}_\ell^d)^T \boldsymbol{M}_{\overline{N}_\ell \overline{N}_\ell}\boldsymbol{y}_\ell^d +\\ + \frac{\alpha}{2}(\boldsymbol{u}_\ell - \boldsymbol{u}_\ell^d)^T \boldsymbol{M}_{\overline{N}_\ell \overline{N}_\ell}(\boldsymbol{u}_\ell - \boldsymbol{u}_\ell^d),\end{aligned}$$

and the optimal control problem (4.2.4) admits the following vector formulation:

$$\begin{aligned}&\textit{minimize} && J_\ell(\boldsymbol{y}_\ell, \boldsymbol{u}_\ell)\\ &\textit{over} && (\boldsymbol{y}_\ell, \boldsymbol{u}_\ell) \in \mathbb{R}^{N_\ell} \times \mathbb{R}^{\overline{N}_\ell},\\ &\textit{subject to} && \boldsymbol{K}_{N_\ell N_\ell}\boldsymbol{y}_\ell = \boldsymbol{M}_{N_\ell \overline{N}_\ell}(\boldsymbol{u}_\ell + \boldsymbol{f}_\ell) - \boldsymbol{\sigma}_\ell, && \text{(4.4.4a)}\\ &&& \boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell \geq \boldsymbol{0}_{N_\ell}, \quad \boldsymbol{\sigma}_\ell \geq \boldsymbol{0}_{N_\ell}, \quad \boldsymbol{\sigma}_\ell \cdot (\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell) = 0. && \text{(4.4.4b)}\end{aligned}$$

This problem is equivalent to MPCC (4.4.1) if we set

$$\begin{aligned}&\boldsymbol{z} := (\boldsymbol{y}_\ell^T, \boldsymbol{u}_\ell^T)^T, \quad f(\boldsymbol{z}) := J_\ell(\boldsymbol{y}_\ell, \boldsymbol{u}_\ell), \quad s := N_\ell + \overline{N}_\ell, \quad m := N_\ell,\\ &F_1(\boldsymbol{z}) := \boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell,\\ &F_2(\boldsymbol{z}) := \boldsymbol{\sigma}_\ell = -\boldsymbol{K}_{N_\ell N_\ell}\boldsymbol{y}_\ell + \boldsymbol{M}_{N_\ell \overline{N}_\ell}(\boldsymbol{u}_\ell + \boldsymbol{f}_\ell),\\ &\boldsymbol{g}(\boldsymbol{z}) := 0, \quad \boldsymbol{h}(\boldsymbol{z}) := 0.\end{aligned}$$

Due to the absence of the inequality constraint (4.4.1b), MFCQ and SMFCQ are identical when applied to problem (4.4.4). Let us now show that (S)MFCQ are satisfied at all feasible points of (4.4.4).

Proposition 4.4.5. *(S)MFCQ holds at any feasible point of* (4.4.4).

Proof. Consider a matrix $A \in \mathbb{R}^{2N_\ell \times (N_\ell + \overline{N}_\ell)}$ constructed of the Jacobians of the vector-valued functions $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$:

$$A = \begin{bmatrix} \nabla_{\boldsymbol{z}} \boldsymbol{F}_1(\boldsymbol{z}) \\ \nabla_{\boldsymbol{z}} \boldsymbol{F}_2(\boldsymbol{z}) \end{bmatrix} = \begin{bmatrix} -\boldsymbol{I}_{N_\ell N_\ell} & \boldsymbol{0}_{N_\ell \overline{N}_\ell} \\ -\boldsymbol{K}_{N_\ell N_\ell} & \boldsymbol{M}_{N_\ell \overline{N}_\ell} \end{bmatrix}.$$

The gradient vectors from Definition 4.4.1 $(i.1)$ and $(ii.1)$ are selected rows of A. Obviously, the rows of A are linearly independent. Further, conditions $(i.2)$ and $(ii.2)$ are satisfied with $\boldsymbol{v} = \boldsymbol{0}_{N_\ell + \overline{N}_\ell}$ (trivial in our case, this condition becomes more involved in presence of inequality constraints). □

Corollary 4.4.6. *Every local minimizer of* (4.4.4) *is simultaneously an S-stationary and a C-stationary point of* (4.4.4).

From the relations (4.4.2) and (4.4.3), we derive the systems characterizing C- and S-stationary points of (4.4.4). Observing that in case of (4.4.4) the Lagrangian functional reduces to

$$L(\boldsymbol{z}, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2) = f(\boldsymbol{z}) - \boldsymbol{F}_1(\boldsymbol{z})^T \boldsymbol{\Gamma}_1 - \boldsymbol{F}_2(\boldsymbol{z})^T \boldsymbol{\Gamma}_2,$$

with a Jacobian

$$\nabla_{\boldsymbol{z}} L(\boldsymbol{z}, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2) = \begin{bmatrix} \boldsymbol{M}_{N_\ell N_\ell} \boldsymbol{y}_\ell - \boldsymbol{M}_{N_\ell \overline{N}_\ell} \boldsymbol{y}_\ell^d + \boldsymbol{\Gamma}_1 + \boldsymbol{K}_{N_\ell N_\ell} \boldsymbol{\Gamma}_2 \\ \alpha \boldsymbol{M}_{\overline{N}_\ell \overline{N}_\ell} (\boldsymbol{u}_\ell - \boldsymbol{u}_\ell^d) - \boldsymbol{M}_{\overline{N}_\ell N_\ell} \boldsymbol{\Gamma}_2 \end{bmatrix}$$

and renaming the vectors of multipliers according to $\boldsymbol{p}_\ell := \boldsymbol{\Gamma}_2 \in \mathbb{R}^{N_\ell}$ and $\boldsymbol{\mu}_\ell := \boldsymbol{\Gamma}_1 \in \mathbb{R}^{N_\ell}$, we obtain the following system that incorporates the feasibility and the optimality conditions, and thus, fully describes S-stationary points of problem (4.4.4).

$$\boldsymbol{K}_{N_\ell N_\ell} \boldsymbol{y}_\ell = \boldsymbol{M}_{N_\ell \overline{N}_\ell} (\boldsymbol{u}_\ell + \boldsymbol{f}_\ell) - \boldsymbol{\sigma}_\ell, \tag{4.4.5a}$$

$$\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell \geq \boldsymbol{0}_{N_\ell}, \quad \boldsymbol{\sigma}_\ell \geq \boldsymbol{0}_{N_\ell}, \quad \boldsymbol{\sigma}_\ell \cdot (\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell) = 0, \tag{4.4.5b}$$

$$\boldsymbol{K}_{N_\ell N_\ell} \boldsymbol{p}_\ell = \boldsymbol{M}_{N_\ell \overline{N}_\ell} \boldsymbol{y}_\ell^d - \boldsymbol{M}_{N_\ell N_\ell} \boldsymbol{y}_\ell - \boldsymbol{\mu}_\ell, \tag{4.4.5c}$$

$$\boldsymbol{p}_\ell = \alpha (\boldsymbol{u}_\ell(\boldsymbol{n}_\ell) - \boldsymbol{u}_\ell^d(\boldsymbol{n}_\ell)), \quad \boldsymbol{u}_\ell(\boldsymbol{n}_\ell^\Gamma) = \boldsymbol{u}_\ell^d(\boldsymbol{n}_\ell^\Gamma), \tag{4.4.5d}$$

$$\boldsymbol{\sigma}_\ell * \boldsymbol{p}_\ell = \boldsymbol{0}_{N_\ell}, \tag{4.4.5e}$$

$$\boldsymbol{\mu}_\ell * (\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell) = \boldsymbol{0}_{N_\ell}, \tag{4.4.5f}$$

$$(\boldsymbol{p}_\ell)_i \geq 0, \ (\boldsymbol{\mu}_\ell)_i \geq 0, \ i : (\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell)_i = (\boldsymbol{\sigma}_\ell)_i = 0. \tag{4.4.5g}$$

A C-stationary point shall, on the other hand, satisfy the conditions (4.4.5a) - (4.4.5f) and

$$(\boldsymbol{p}_\ell)_i (\boldsymbol{\mu}_\ell)_i \geq 0, \ i : (\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell)_i = (\boldsymbol{\sigma}_\ell)_i = 0. \tag{4.4.5h}$$

Remark 4.4.7. Obviously, every S-stationary point is also a C-stationary point of (4.4.4).

4.4.3 Stationarity concepts for problem (4.2.4)

Finally, taking into account (4.2.1), we can reformulate the system (4.4.5) in terms of finite element functions. By that, we provide definitions of S-stationary and C-stationary points for the discrete optimal control problem (4.2.4).

Note that S- and C-stationary points differ only by the properties of p_ℓ and μ_ℓ in the biactive nodes. If strict complementarity is satisfied at a local minimizer of (4.2.4) (i.e., $B_\ell = \varnothing$), S- and C-stationarity systems reduce to one and the same system of conditions.

Definition 4.4.8. *(S-stationary, C-stationary, and stationary points with strict complementarity)*
Let $(y_\ell, \sigma_\ell, u_\ell, p_\ell, \mu_\ell) \in V_\ell \times \mathcal{M}_\ell \times S_\ell \times V_\ell \times \mathcal{M}_\ell$ be given and satisfy

$$a(y_\ell, v_\ell) = (f + u_\ell, v_\ell)_{0,\Omega} - \langle\langle \sigma_\ell, v_\ell \rangle\rangle, \ \forall\, v_\ell \in V_\ell, \tag{4.4.6a}$$

$$y_\ell - \psi_\ell \leq 0, \ \sigma_\ell \in \mathcal{M}_\ell \cap \mathcal{M}_+(\overline{\Omega}), \ \langle\langle \sigma_\ell, y_\ell - \psi_\ell \rangle\rangle = 0, \tag{4.4.6b}$$

$$a(p_\ell, v_\ell) = (y^d - y_\ell, v_\ell)_{0,\Omega} - \langle\langle \mu_\ell, v_\ell \rangle\rangle, \ \forall\, v_\ell \in V_\ell, \tag{4.4.6c}$$

$$p_\ell = \alpha(u_\ell - u_\ell^d), \tag{4.4.6d}$$

$$p_\ell(a) = 0, \ a \in C_\ell, \tag{4.4.6e}$$

$$\mu_\ell(a) = 0, \ a \in I_\ell. \tag{4.4.6f}$$

(i) The triple $(y_\ell, \sigma_\ell, u_\ell) \in V_\ell \times \mathcal{M}_\ell \times S_\ell$ is called an S-stationary point, if the pair $(p_\ell, \mu_\ell) \in V_\ell \times \mathcal{M}_\ell$ fulfills

$$\mu_\ell(a) \geq 0, \ p_\ell(a) \geq 0, \ a \in B_\ell. \tag{4.4.6g}$$

(ii) The triple $(y_\ell, \sigma_\ell, u_\ell) \in V_\ell \times \mathcal{M}_\ell \times S_\ell$ is called a C-stationary point, if the pair $(p_\ell, \mu_\ell) \in V_\ell \times \mathcal{M}_\ell$ fulfills

$$\mu_\ell(a) p_\ell(a) \geq 0, \ a \in B_\ell. \tag{4.4.6h}$$

(iii) The triple $(y_\ell, \sigma_\ell, u_\ell) \in V_\ell \times \mathcal{M}_\ell \times S_\ell$ is called a stationary point with strict complementarity, if $B_\ell = \varnothing$, i.e.,

$$C_\ell = A_\ell. \tag{4.4.6i}$$

The function p_ℓ is referred to as the discrete adjoint state, the equations (4.4.6a) *and* (4.4.6c) *as the discrete state and the discreet adjoint state equations, respectively. The functionals σ_ℓ and μ_ℓ, are referred to as the discrete Lagrangian multipliers associated with the discrete state and the discrete adjoint state equations, respectively.*

Remark 4.4.9. Due to (4.4.6e) and (4.4.6f), condition (4.4.6h) implies

$$\langle\langle \mu_\ell, p_\ell \rangle\rangle \geq 0. \tag{4.4.7}$$

Note that the reverse is not true. $\langle\langle \mu_\ell, p_\ell \rangle\rangle = \sum_{a \in B_\ell} \mu_\ell(a) p_\ell(a) \geq 0$ does not imply that every summand is nonnegative. I.e., condition (4.4.7) is weaker then (4.4.6h). The implications of this observation will be addressed in the next paragraph.

Remark 4.4.10. Corollary 4.4.6 implies that problem (4.2.4) always has an S-stationary point (and, consequently, a C-stationary point). When solving the problem numerically, it makes sense to look for points satisfying the system (4.4.6a)-(4.4.6f), (4.4.6g) - the strongest set of necessary optimality conditions we have at hand. Note that unlike the systems of S- and C-stationarity conditions, there is no guarantee that the system (4.4.6a)-(4.4.6f), (4.4.6i) will have a solution for any given set of data. For some problems the strict complementarity condition might be violated at all feasible points.

Remark 4.4.11. By the Definition 2.3.5, when μ_ℓ is considered as a functional on V_ℓ, there holds

$$\mathcal{O}_{\mu_\ell} \supseteq \operatorname{int}(\mathcal{I}_\ell), \quad \operatorname{supp}(\mu_\ell) \subseteq \mathcal{A}_\ell,$$

whereas, if understood as an element of $\mathcal{M}(\overline{\Omega})$, it satisfies

$$\mathcal{O}_{\mu_\ell} \supseteq I_\ell, \quad \operatorname{supp}(\mu_\ell) \subseteq A_\ell.$$

The inclusion is due to the fact that neither of the stationarity systems forbids the coefficients of μ_ℓ to be non-zero in the active nodes (cf. Remark 4.3.15).

4.4.4 Continuous analogs of the local properties of the discrete Lagrangian multipliers

Now that the stationarity systems for the continuous and the discrete optimal control problems have been introduced, we can compare how the localization of the Lagrangian multipliers is carried out in the functional spaces and in the finite dimensional setting.

First, let us regard the basic operations often used in the discrete setting - scalar and Hadamard products of vectors - and, when available, provide their analogs in the functional space setting. Only within this paragraph, we allow the pair (λ, v) to have three different qualities:

- $(\lambda, v) = (\boldsymbol{\lambda}, \boldsymbol{v}) \in \mathbb{R}^{N_\ell} \times \mathbb{R}^{N_\ell}$ ($\mathbb{R}^{N_\ell}$-setting),
- $(\lambda, v) \in L^2(\Omega) \times L^2(\Omega)$ ($L^2(\Omega)$-setting),
- $(\lambda, v) \in V^* \times V$ (V^*-setting).

The table below summarizes the known equivalents of the scalar and the Hadamard products of vectors for the infinite-dimensional cases $(\lambda, v) \in L^2(\Omega) \times L^2(\Omega)$ and $(\lambda, v) \in V^* \times V$:

	$\lambda,\ v \in \mathbb{R}^{N_\ell}$	$\lambda,\ v \in L^2(\Omega)$	$\lambda \in V^*,\ v \in V$
scalar product	$\lambda \cdot v$	$(\lambda, v)_{0,\Omega}$	$\langle \lambda, v \rangle$
Hadamard product	$\lambda * v$	λv	—

Both operations are available in the $L^2(\Omega)$-setting. In the V^*-setting, however, due to the fact that $\lambda \in V^*$ does not in general have a pointwise interpretation, no obvious equivalent of the Hadamard product is available. Nevertheless, some properties formulated (solely or in part) in terms of the Hadamard product in $\mathbb{R}^n$ (point-wise a.e. product in $L^2(\Omega)$) can be equivalently written in terms of scalar products in $\mathbb{R}^n$ ($L^2(\Omega)$). In this way, they gain an interpretation in the V^*-setting. Some examples of such properties are shown in the following table. For $v \in \mathbb{R}^n$, $v^+ := \max(\mathbf{0}_n, v)$ and

$\lambda,\, v \in \mathbb{R}^{N_\ell}$	$\lambda,\, v \in L^2(\Omega)$	$\lambda \in V^*,\, v \in V$
$\lambda * v = \mathbf{0}_n$	$\lambda v = 0$ a.e. in Ω	—
$\Updownarrow$	$\Updownarrow$	
$\lambda \cdot (v * \varphi) = \mathbf{0}_n,\ \forall\, \varphi \in \mathbb{R}^n$	$(\lambda, v\varphi)_{0,\Omega} = 0,\ \forall\, \varphi \in C^1(\overline{\Omega})$	$\langle \lambda, v\varphi \rangle = 0,\ \forall\, \varphi \in C^1(\overline{\Omega})$
$\lambda \geq \mathbf{0}_n$ and	$\lambda \in (L^2(\Omega))_+$ and	$\lambda \in V_+^*$
$\lambda * v = \mathbf{0}_n$	$\lambda v = 0$ a.e. in Ω	—
$\Updownarrow$	$\Updownarrow$	
$\lambda \geq \mathbf{0}_n$ and	$\lambda \in (L^2(\Omega))_+$ and	$\lambda \in V_+^*$ and
$\lambda \cdot v^+ = \lambda \cdot v^- = \mathbf{0}_n$	$(\lambda, v^+)_{0,\Omega} = (\lambda, v^-)_{0,\Omega} = 0$	$\langle \lambda, v^+ \rangle = \langle \lambda, v^- \rangle = 0$
$\lambda * v \geq \mathbf{0}_n$	$\lambda v \geq 0$ a.e. in Ω	—
$\Updownarrow$	$\Updownarrow$	
$\lambda \cdot (v * \varphi) \geq \mathbf{0}_n,$	$(\lambda, v\varphi)_{0,\Omega} \geq 0,$	$\langle \lambda, v\varphi \rangle \geq 0,$
$\forall\, \varphi \in \mathbb{R}^n,\, \varphi \geq \mathbf{0}_n$	$\forall\, \varphi \in C^1(\overline{\Omega}) \cap (L^2(\Omega))_+$	$\forall\, \varphi \in C^1(\overline{\Omega}) \cap V_+$

Table 4.1: Table of correspondences between selected properties of dual pairings in $\mathbb{R}^n$, $L^2(\Omega)$, and V^*.

$v^- := \min(\mathbf{0}_n, v)$ with "max" and "min" understood componentwise.

The rest of this section is devoted to selected parts of the discrete stationarity systems which involve the scalar and the Hadamard products. Using the rules introduced above, continuous analogs of these properties are suggested for the cases $(y, \sigma) \in V \cap H^2(\Omega) \times L^2(\Omega)$ and $(y, \sigma) \in V \times V^*$ and, if detected, the disagreements with Definitions 3.3.3 and 3.3.8 are pointed out.

Remark 4.4.12. The V^*-properties enlisted in Table 4.1 are only suggestions of possible realizations of corresponding properties in the $\mathbb{R}^{N_\ell}$- and $L^2(\Omega)$-settings.

Localization by means of the scalar product

In the discrete system, the complementarity property of the non-negative vectors $\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell$ and $\boldsymbol{\sigma}_\ell$ (4.4.5b) is expressed in terms of a scalar product:

$$\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell \geq \mathbf{0}_{N_\ell}, \quad \boldsymbol{\sigma}_\ell \geq \mathbf{0}_{N_\ell}, \quad \boldsymbol{\sigma}_\ell \cdot (\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell) = 0. \tag{4.4.8}$$

An obvious equivalent for the $L^2(\Omega)$-setting is given by

$$\psi - y \in V_+, \ \sigma \in (L^2(\Omega))_+, \ (\sigma, y - \psi)_{0,\Omega} = 0,$$

(cf. (3.3.2b)), whereas for $\sigma \in V^*$ the variant employed in the system (3.3.11) is available:

$$\psi - y \in V_+, \ \sigma \in V_+^*, \ \langle \sigma, y - \psi \rangle = 0.$$

Localization by means of the Hadamard product

Localization in the strongly active set. In the finite-dimensional setting, the structural dependence between $\boldsymbol{\sigma}_\ell$ and $\boldsymbol{p}_\ell$ is expressed by the Hadamard product

$$\boldsymbol{\sigma}_\ell * \boldsymbol{p}_\ell = \mathbf{0}_{N_\ell} \tag{4.4.9}$$

(cf. (4.4.5e)). Note that, independently on the signs of $\boldsymbol{\sigma}_\ell$ and $\boldsymbol{p}_\ell$, this condition enforces a complementarity relation between these two vectors (i.e., their components at matching positions cannot be equal to zero simultaneously). Thus, for instance, $\boldsymbol{\sigma}_\ell(\boldsymbol{n}_\ell^C) \neq \mathbf{0}_{N_\ell^C}$ in combination with (4.4.9) implies $\boldsymbol{p}_\ell(\boldsymbol{n}_\ell^C) = \mathbf{0}_{N_\ell^C}$.

Taking into account the correspondences in Table 4.1, for $\sigma \in L^2(\Omega)$ we could require

$$\sigma p = 0 \text{ a.e. in } \Omega, \tag{4.4.10}$$

and for $\sigma \in V^*$,

$$\langle \sigma, vp \rangle = 0, \ \forall\, \varphi \in C^1(\overline{\Omega}). \tag{4.4.11}$$

On the other hand, in view of the non-negativity of $\boldsymbol{\sigma}_\ell$ in the finite-dimensional setting, as well as of σ in the infinite-dimensional settings with $\sigma \in L^2(\Omega)$ and $\sigma \in V^*$, we could employ the realizations

$$\boldsymbol{\sigma}_\ell \cdot \boldsymbol{p}_\ell^+ = \boldsymbol{\sigma}_\ell \cdot \boldsymbol{p}_\ell^- = \mathbf{0}_{N_\ell}, \tag{4.4.12}$$

$$(\sigma, p^+)_{0,\Omega} = (\sigma, p^-)_{0,\Omega} = 0,$$

and

$$\langle \sigma, p^+ \rangle = \langle \sigma, p^- \rangle = 0$$

instead of (4.4.9), (4.4.10), and (4.4.11), respectively.

Localization in the biactive set. Discrete C-stationarity points satisfy

$$\boldsymbol{p}_\ell(\boldsymbol{n}_\ell^B) * \boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell^B) \geq \mathbf{0}_{N_\ell^B}, \tag{4.4.13}$$

(cf. (4.4.6h)). Due to $\boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell^I) = \mathbf{0}_{N_\ell^I}$ and $\boldsymbol{p}_\ell(\boldsymbol{n}_\ell^C) = \mathbf{0}_{N_\ell^C}$, this condition is equivalent to

$$\boldsymbol{p}_\ell(\boldsymbol{n}_\ell) * \boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell) \geq \mathbf{0}_{N_\ell}. \tag{4.4.14}$$

For the adjoint state $p \in V$ and the Lagrangian multiplier associated with the adjoint state equation $\mu \in V^*$ in the infinite-dimensional setting, condition (4.4.14) can be rewritten as

$$\langle \mu, p\varphi \rangle \geq 0 \; \forall \, \varphi \in C^1(\overline{\Omega}) \cap V_+. \tag{4.4.15}$$

Remark 4.4.13. It should be noted that in the continuous versions of the C-stationarity concept introduced in Definitions 3.3.3 and 3.3.8 the pairing of μ and p is subject to a weaker requirement:

$$\langle \mu, p \rangle \geq 0. \tag{4.4.16}$$

Obviously, (4.4.16) is a special case of (4.4.15) with $\varphi \equiv 1$. Moreover, (4.4.16) rewritten for the $\mathbb{R}^n$-setting reads $\boldsymbol{p}_\ell(\boldsymbol{n}_\ell) \cdot \boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell) \geq \mathbf{0}_{N_\ell}$, which, in turn, is obviously weaker than (4.4.14) (cf. Remark 4.4.9). Thus, in the aspect of localization of the adjoint state and the Lagrangian multiplier associated with the adjoint state in the biactive set, the continuous C-stationarity concepts of Definitions 3.3.3 and 3.3.8 are less restrictive than the well-established concept of C-stationarity for the finite-dimensional MPEC.

4.5 Extension of the discrete Lagrangian multipliers

In the finite dimensional setting, the operation of the discrete Lagrangian multipliers σ_ℓ and μ_ℓ on functions $v_\ell \in V_\ell$ can be expressed in terms of the more regular solution components. This observation allows to construct extensions $\hat{\sigma}_\ell$, $\hat{\mu}_\ell$ and $\tilde{\sigma}_\ell$, $\tilde{\mu}_\ell$ to functionals on V that provide more natural approximations to σ, $\mu \in V^*$. The extensions $\hat{\sigma}_\ell$, $\hat{\mu}_\ell$ will be used in the convergence analysis for the finite element approximations in Chapter 5, whereas $\tilde{\sigma}_\ell$, $\tilde{\mu}_\ell$ will play an essential role in the a posteriori error analysis in Chapter 6.

Let us first introduce the following auxiliary operator.

Definition 4.5.1. *For any subset of interior nodes $D_\ell \subseteq \mathcal{N}_\ell$, the operator $I_{D_\ell} : V_\ell \to V_\ell$ is defined according to*

$$I_{D_\ell}(v_\ell)(a) := \begin{cases} v_\ell(a), & a \in D_\ell \\ 0 \quad\;\; , & a \in \mathcal{N}_\ell \setminus D_\ell \end{cases}, \quad \forall \; v_\ell \in V_\ell \, .$$

There obviously holds

$$I_{C_\ell}(v_\ell)|_{\mathcal{C}_\ell} = v_\ell|_{\mathcal{C}_\ell}, \quad I_{C_\ell}(v_\ell)|_{\overset{\circ}{\mathcal{Z}}_\ell} = 0, \tag{4.5.1a}$$

$$I_{C_\ell}(v_\ell)|_D = \sum_{a \in \mathcal{N}_\ell(D) \cap C_\ell} v_\ell(a) \varphi_\ell^{(a)}, \; \forall \, D \in \{T, E\}, \; T \subseteq \mathcal{F}_\ell(\sigma_\ell), \; E \in \mathcal{E}_{\mathcal{F}_\ell(\sigma)}, \tag{4.5.1b}$$

and

$$I_{A_\ell}(v_\ell)|_{\mathcal{A}_\ell} = v_\ell|_{\mathcal{A}_\ell}, \quad I_{A_\ell}(v_\ell)|_{\overset{\circ}{\mathcal{I}}_\ell} = 0, \tag{4.5.1c}$$

$$I_{A_\ell}(v_\ell)|_D = \sum_{a\in\mathcal{N}_\ell(D)\cap A_\ell} v_\ell(a)\varphi_\ell^{(a)}, \ \forall\, D\in\{T,E\},\ T\subseteq\mathcal{F}_\ell(y_\ell),\ E\in\mathcal{E}_{\mathcal{F}_\ell(y_\ell)}. \tag{4.5.1d}$$

It is easy to see that the action of the discrete multipliers on elements of V_ℓ complies with

$$\langle\langle\sigma_\ell, v_\ell\rangle\rangle = \langle\langle\sigma_\ell, I_{C_\ell}(v_\ell)\rangle\rangle, \quad \langle\langle\mu_\ell, v_\ell\rangle\rangle = \langle\langle\mu_\ell, I_{A_\ell}(v_\ell)\rangle\rangle, \quad \forall\, v_\ell\in V_\ell.$$

A more detailed examination yields the following representation of the discrete Lagrangian multipliers.

Proposition 4.5.2. *The action of σ_ℓ and μ_ℓ on $v_\ell\in V_\ell$ is determined by:*

$$\begin{aligned}\langle\langle\sigma_\ell, v_\ell\rangle\rangle &= \sum_{T\subseteq\mathcal{C}_\ell\cup\mathcal{F}_\ell(\sigma_\ell)} [(f+u_\ell, I_{C_\ell}(v_\ell))_{0,T} - (\nabla y_\ell, \nabla I_{C_\ell}(v_\ell))_{0,T}] \qquad (4.5.2a)\\ &= \sum_{T\subseteq\mathcal{C}_\ell\cup\mathcal{F}_\ell(\sigma_\ell)} (f+u_\ell, I_{C_\ell}(v_\ell))_{0,T} - \sum_{E\in\mathcal{E}_{\mathcal{C}_\ell}\cup\mathcal{E}_{\mathcal{F}_\ell(\sigma_\ell)}} (\nu_E\cdot[\nabla y_\ell], I_{C_\ell}(v_\ell))_{0,E},\\ \langle\langle\mu_\ell, v_\ell\rangle\rangle &= \sum_{T\subseteq\mathcal{A}_\ell\cup\mathcal{F}_\ell(y_\ell)} \Big[(y^d-y_\ell, I_{A_\ell}(v_\ell))_{0,T} - (\nabla p_\ell, \nabla I_{A_\ell}(v_\ell))_{0,T}\Big] \qquad (4.5.2b)\\ &= \sum_{T\subseteq\mathcal{A}_\ell\cup\mathcal{F}_\ell(y_\ell)} (y^d-y_\ell, I_{A_\ell}(v_\ell))_{0,T} - \sum_{E\in\mathcal{E}_{\mathcal{A}_\ell}\cup\mathcal{E}_{\mathcal{F}_\ell(y_\ell)}} (\nu_E\cdot[\nabla p_\ell], I_{A_\ell}(v_\ell))_{0,E}.\end{aligned}$$

Proof. The nodal coefficients of the discrete multipliers σ_ℓ and μ_ℓ admit the representation

$$\sigma_\ell(a) = \begin{cases} \sum_{T\in\omega_\ell^a}\Big[(f+u_\ell, \varphi_\ell^{(a)})_{0,T} - (\nabla y_\ell, \nabla\varphi_\ell^{(a)})_{0,T}\Big], & a\in C_\ell \\ 0, & a\in Z_\ell \end{cases}, \tag{4.5.3a}$$

$$\mu_\ell(a) = \begin{cases} \sum_{T\in\omega_\ell^a}\Big[(y^d-y_\ell, \varphi_\ell^{(a)})_{0,T} - (\nabla p_\ell, \nabla\varphi_\ell^{(a)})_{0,T}\Big], & a\in A_\ell \\ 0, & a\in I_\ell \end{cases}, \tag{4.5.3b}$$

as well as

$$\sigma_\ell(a) = \begin{cases} \sum_{T\in\omega_\ell^a}(f+u_\ell, \varphi_\ell^{(a)})_{0,T} - \sum_{E\in\mathcal{E}_\ell^a}(\nu_E\cdot[\nabla y_\ell], \varphi_\ell^{(a)})_{0,E}, & a\in C_\ell \\ 0, & a\in Z_\ell \end{cases}, \tag{4.5.4a}$$

$$\mu_\ell(a) = \begin{cases} \sum_{T\in\omega_\ell^a}(y^d-y_\ell, \varphi_\ell^{(a)})_{0,T} - \sum_{E\in\mathcal{E}_\ell^a}(\nu_E\cdot[\nabla p_\ell], \varphi_\ell^{(a)})_{0,E}, & a\in A_\ell \\ 0, & a\in I_\ell \end{cases}. \tag{4.5.4b}$$

Indeed, (4.4.6f) and Definition 4.3.9 yield $\sigma_\ell(a) = 0$, $a \in Z_\ell$ and $\mu_\ell(a) = 0$, $a \in I_\ell$. Applying σ_ℓ and μ_ℓ to $v_\ell = \varphi_\ell^{(a)}$ in view of (4.4.6a) and (4.4.6c) we obtain

$$\langle\langle \sigma_\ell, \varphi_\ell^{(a)} \rangle\rangle = (f + u_\ell, \varphi_\ell^{(a)})_{0,\omega_\ell^a} \; - \; (\nabla y_\ell, \nabla \varphi_\ell^{(a)})_{0,\omega_\ell^a}\,, \; a \in C_\ell\,,$$
$$\langle\langle \mu_\ell, \varphi_\ell^{(a)} \rangle\rangle = (y^d - y_\ell, \varphi_\ell^{(a)})_{0,\omega_\ell^a} \; - \; (\nabla p_\ell, \nabla \varphi_\ell^{(a)})_{0,\omega_\ell^a}\,, \; a \in A_\ell\,,$$

which proves (4.5.3). Representation (4.5.4) can be obtained by element-wise application of Green's formula to the second terms on the right-hand side of the latter expressions.
Taking $\langle\langle \sigma_\ell, v_\ell \rangle\rangle = \sum\limits_{a \in \mathcal{N}_\ell} \sigma_\ell(a) v_\ell(a) = \sum\limits_{a \in C_\ell} \sigma_\ell(a) v_\ell(a)$ into account, from(4.5.3) and (4.5.4) we deduce

$$\langle\langle \sigma_\ell, v_\ell \rangle\rangle = \sum_{a \in C_\ell} \left(\sum_{T \in \omega_\ell^a} \left[(f + u_\ell, v_\ell(a) \varphi_\ell^{(a)})_{0,T} - (\nabla y_\ell, \nabla v_\ell(a) \varphi_\ell^{(a)})_{0,T} \right] \right),$$

and

$$\langle\langle \sigma_\ell, v_\ell \rangle\rangle = \sum_{a \in C_\ell} \left(\sum_{T \in \omega_\ell^a} (f + u_\ell, v_\ell(a) \varphi_\ell^{(a)})_{0,T} - \sum_{E \in \mathcal{E}_\ell^a} (\nu_E \cdot [\nabla y_\ell], v_\ell(a) \varphi_\ell^{(a)})_{0,E} \right),$$

respectively. Regrouping the summands in the above expressions and taking into account (4.5.1), we deduce the assertion. The proof for μ_ℓ follows the same lines. □

Let us now introduce extensions of the discrete Lagrangian multipliers to V^*. We call $E_\ell^\lambda \in V^*$ an extension of $\lambda_\ell \in \mathcal{M}_\ell$ to V^* if the operation of E_ℓ^λ on V_ℓ is equal to that of λ_ℓ, i.e.,

$$\langle E_\ell^\lambda, v_\ell \rangle = \langle\langle \lambda_\ell, v_\ell \rangle\rangle, \; \forall\, v_\ell \in V_\ell.$$

We restrict ourselves to two types of extensions of the discrete Lagrangian multipliers. The first extensions $\hat{\sigma}_\ell$, $\hat{\mu}_\ell \in V^*$ are defined in much the same way as in the finite element analysis of variational inequalities of obstacle type (cf., e.g., [11]), whereas the second extensions $\tilde{\sigma}_\ell$, $\tilde{\mu}_\ell \in V^*$ are defined in view of Proposition 4.5.2.

Definition 4.5.3. *(Extensions of the discrete Lagrangian multipliers)*
We define the functionals $\hat{\sigma}_\ell$, $\hat{\mu}_\ell \in V^*$ by means of

$$\langle \hat{\sigma}_\ell, v \rangle := (f + u_\ell, v)_{0,\Omega} - a(y_\ell, v), \; \forall\, v \in V, \tag{4.5.5a}$$
$$\langle \hat{\mu}_\ell, v \rangle := (y^d - y_\ell, v)_{0,\Omega} - a(p_\ell, v), \; \forall\, v \in V, \tag{4.5.5b}$$

and functionals $\tilde{\sigma}_\ell$, $\tilde{\mu}_\ell \in V^*$ according to

$$\langle \tilde{\sigma}_\ell, v \rangle := \sum_{T \subseteq \mathcal{C}_\ell} (f + u_\ell, v)_{0,T} - \sum_{E \in \mathcal{E}_{\mathcal{C}_\ell}} (\nu_E \cdot [\nabla y_\ell], v)_{0,E} + F_{\ell,\sigma}(P_{\ell,SZ} v), \; \forall\, v \in V, \tag{4.5.6a}$$
$$\langle \tilde{\mu}_\ell, v \rangle := \sum_{T \subseteq \mathcal{A}_\ell} (y^d - y_\ell, v)_{0,T} - \sum_{E \in \mathcal{E}_{\mathcal{A}_\ell}} (\nu_E \cdot [\nabla p_\ell], v)_{0,E} + F_{\ell,\mu}(P_{\ell,SZ} v), \; \forall\, v \in V, \tag{4.5.6b}$$

where $P_{\ell,SZ} : V \to V_\ell$ is the Scott-Zhang interpolation operator and

$$F_{\ell,\sigma}(v_\ell) := \sum_{T\subseteq\mathcal{F}_\ell(\sigma_\ell)} (f + u_\ell, I_{C_\ell}(v_\ell))_{0,T} - \sum_{E\in\mathcal{E}_{\mathcal{F}_\ell(\sigma_\ell)}} (\nu_E \cdot [\nabla y_\ell], I_{C_\ell}(v_\ell))_{0,E},$$
$$F_{\ell,\mu}(v_\ell) := \sum_{T\subseteq\mathcal{F}_\ell(y_\ell)} (y^d - y_\ell, I_{A_\ell}(v_\ell))_{0,T} - \sum_{E\in\mathcal{E}_{\mathcal{F}_\ell(y_\ell)}} (\nu_E \cdot [\nabla p_\ell], I_{A_\ell}(v_\ell))_{0,E}.$$

In the second form of the extensions, the use of the interpolation operator is essential as the functionals $F_{\ell,\mu}$ and $F_{\ell,\sigma}$ can only be evaluated on elements of V_ℓ.

Proposition 4.5.4. *The functionals $\hat\sigma_\ell$, $\hat\mu_\ell \in V^*$ and $\tilde\sigma_\ell$, $\tilde\mu_\ell \in V^*$ are extensions of σ_ℓ, $\mu_\ell \in \mathcal{M}_\ell$, i.e., for $v_\ell \in V_\ell$ there holds*

$$\langle\hat\sigma_\ell, v_\ell\rangle = \langle\tilde\sigma_\ell, v_\ell\rangle = \langle\langle\sigma_\ell, v_\ell\rangle\rangle,$$
$$\langle\hat\mu_\ell, v_\ell\rangle = \langle\tilde\mu_\ell, v_\ell\rangle = \langle\langle\mu_\ell, v_\ell\rangle\rangle.$$

Proof. The results are immediate consequences of (4.4.6) and Proposition 4.5.2. □

Remark 4.5.5.

(*i*) Though the two forms of the extensions coincide on V_ℓ, on the whole space V they are not the same. The difference lies essentially along the discrete free boundaries $\mathcal{F}_\ell(y_\ell)$ and $\mathcal{F}_\ell(\sigma_\ell)$.

(*ii*) It is clear that $\hat\sigma_\ell$, $\tilde\sigma_\ell \notin V_+^*$. The same can be said about the sign properties of $\hat\mu_\ell$, $\tilde\mu_\ell$ in the discrete biactive set in case of S-stationary points. Thus, neither of the two extensions constitutes a conforming approximation for the Lagrangian multipliers σ, $\mu \in V^*$.

(*iii*) Due to Proposition 4.5.4, the extensions inherit the properties of σ_ℓ and μ_ℓ with respect to the elements of V_ℓ such as (4.4.6b) and (4.4.7), as well as all the properties that follow from the local characterizations (4.4.6f), (4.4.6g), and (4.4.6h). Also, if viewed as functionals on V_ℓ, the extensions fulfill

$$\operatorname{supp}(\hat\sigma_\ell) = \operatorname{supp}(\tilde\sigma_\ell) = \operatorname{supp}(\sigma_\ell) = C_\ell, \quad \operatorname{supp}(\hat\mu_\ell) = \operatorname{supp}(\tilde\mu_\ell) = \operatorname{supp}(\mu_\ell) \subseteq A_\ell.$$

(*iv*) Local properties of the first extensions $\hat\sigma_\ell$, $\hat\mu_\ell \in V^*$, in particular any information regarding their supports and zero sets, are difficult to obtain, whereas the extensions $\tilde\sigma_\ell$, $\tilde\mu_\ell \in V^*$ obviously satisfy

$$\mathcal{C}_\ell \subseteq \operatorname{supp}(\tilde\sigma_\ell) \subseteq \mathcal{C}_\ell \cup \mathcal{F}_\ell(\sigma_\ell), \tag{4.5.7a}$$
$$\operatorname{supp}(\tilde\mu_\ell) \subseteq \mathcal{A}_\ell \cup \mathcal{F}_\ell(y_\ell) \tag{4.5.7b}$$

(the support of $\tilde\mu_\ell$ might not fill $\mathcal{A}_\ell$, i.e., $\operatorname{supp}(\tilde\mu_\ell) \cap \mathcal{A}_\ell \subseteq \mathcal{A}_\ell$, cf. Remark 4.4.11). The precise structure of $\operatorname{supp}(\tilde\sigma_\ell)$ depends on the definition of the Scott-Zhang interpolation operator $P_{\ell,SZ}$. In particular, under the condition

$$\forall\, a \in C_\ell,\ \exists\, T^{(a)} \subset \omega_\ell^a \ :\ T^{(a)} \subseteq \mathcal{C}_\ell \tag{4.5.8}$$

we obtain $\operatorname{supp}(\tilde{\sigma}_\ell) = \mathcal{C}_\ell$, if the triangles satisfying (4.5.8) are used in the definition of $P_{\ell,SZ}$. Note that (4.5.8) excludes isolated strongly active nodal points and edges (cf. Definition 4.3.12). However, utilizing a Scott-Zhang interpolation operator defined by averaging over edges instead of triangles (see [57]), allows to show $\operatorname{supp}(\tilde{\sigma}_\ell) = \mathcal{C}_\ell$, if we only exclude isolated strongly active nodal points. Similar remarks apply to $\tilde{\mu}_\ell$, i.e., it is possible to achieve $\operatorname{supp}(\tilde{\mu}_\ell) \subseteq \mathcal{A}_\ell$ instead of (4.5.7b), if no isolated active nodal points occur and the modified $P_{\ell,SZ}$ is used. Yet, in general, we cannot avoid the presence of the isolated active and strongly active nodal points.

5 Convergence of the finite element scheme

This section is devoted to the study of the behavior of the sequences of finite element approximations of stationary points $\{(y_\ell, u_\ell, \sigma_\ell)\}_{\mathbb{N}}$ computed on nested triangulations $\mathcal{T}_\ell$, $\ell \in \mathbb{N}$, with $\ell \to \infty$.

5.1 Convergence result

Let us first introduce some supporting definitions and assumptions.

Definition 5.1.1. *The sequence $\{V_\ell\}_{\mathbb{N}}$ of subspaces of V is limit dense in V if for any $v \in V$ there exists a sequence $\{v_\ell\}_{\mathbb{N}}$, $v_\ell \in V_\ell$, $\ell \in \mathbb{N}$, such that $v_\ell \to v$ in V for $\ell \to \infty$.*

As an example of such a sequence we may consider nested finite element subspaces corresponding to a sequence of uniformly refined triangulations $\{\mathcal{T}_\ell\}_{\mathbb{N}}$.

Remark 5.1.2. If $\{V_\ell\}_{\mathbb{N}}$ is limit dense in V, then $\{(V_\ell)_+\}_{\mathbb{N}}$ is limit dense in V_+, where $(V_\ell)_+ := V_\ell \cap V_+$, $\ell \in \mathbb{N}$. The reader is referred to [28], Section 4.3.

Suppose that the following assumptions with respect to the sequences of stationary points $\{(y_\ell, u_\ell, \sigma_\ell)\}_{\mathbb{N}}$ and the data function ψ are satisfied:

Assumption 5.1.3.

(*i*) *$\{(y_\ell, u_\ell, \sigma_\ell)\}_{\mathbb{N}}$ is a sequence of global solutions of* (4.2.2) *or the sequences $\{y_\ell\}_{\mathbb{N}}$ and $\{u_\ell\}_{\mathbb{N}}$ are uniformly bounded in $L^2(\Omega)$, i.e., there exist positive constants C_y, C_u such that*

$$\|y_\ell\|_{0,\Omega} \le C_y, \ \|u_\ell\|_{0,\Omega} \le C_u, \quad \ell \in \mathbb{N};$$

(*ii*) *the obstacle function satisfies $A\psi \in L^2(\Omega)$; with this assumption we may restrict ourselves to $\psi = 0$, since otherwise in* (3.2.1) *we can replace f by $f - A\psi$ and y^d by $y^d - \psi$.*

Theorem 5.1.4. *Let $\{(y_\ell, \sigma_\ell, u_\ell)\}_{\mathbb{N}}$, $(y_\ell, \sigma_\ell, u_\ell) \in V_\ell \times \mathcal{M}_\ell \times S_\ell$, $\ell \in \mathbb{N}$, be a sequence of C-stationary points of* (4.2.2) *with corresponding adjoint states and multipliers $\{(p_\ell, \mu_\ell)\}_{\mathbb{N}}$, $(p_\ell, \mu_\ell) \in V_\ell \times \mathcal{M}_\ell$, $\ell \in \mathbb{N}$ computed on a sequence of nested finite element subspaces $\{V_\ell\}_{\mathbb{N}}$. Let $\hat{\sigma}_\ell \in V^*$ and $\hat{\mu}_\ell \in V^*$ be the extensions of σ_ℓ and μ_ℓ, respectively, as given by* (4.5.5a), (4.5.5b).

If Assumption 5.1.3 is satisfied and the sequence $\{V_\ell\}_\mathbb{N}$ is limit dense in V, then there exist a subsequence $\mathbb{N}' \subset \mathbb{N}$ and an almost C-stationary point of (3.1.1)*,*

$$(y^*, \sigma^*, u^*, p^*, \mu^*) \in V \times V^* \times L^2(\Omega) \times V \times V^*,$$

such that for $\ell \in \mathbb{N}'$, $\ell \to \infty$

$$\begin{aligned}
&y_\ell \to y^* \text{ in } V, \quad y_\ell \to y^* \text{ in } L^2(\Omega),\\
&\hat{\sigma}_\ell \to \sigma^* \text{ in } V^*,\\
&u_\ell \to u^* \text{ in } L^2(\Omega),\\
&p_\ell \rightharpoonup p^* \text{ in } V, \quad p_\ell \to p^* \text{ in } L^2(\Omega),\\
&\hat{\mu}_\ell \rightharpoonup^* \mu^* \text{ in } V^*.
\end{aligned}$$

If $\{S_\ell\}_\mathbb{N}$ is limit dense in $H^1(\Omega)$, then $(y^*, \sigma^*, u^*, p^*, \mu^*)$ satisfies

$$\langle \mu^*, y^* v \rangle = 0, \ \forall\, v \in C^1(\overline{\Omega}). \tag{5.1.1}$$

5.2 Proof of the convergence theorem

In order to prove Theorem 5.1.4, we will first show the convergence properties of the sequence of the discrete solutions and the properties of the limit common for all types of stationary points. In the end, the specific conditions characterizing almost C-stationarity will be addressed.

Proof. Suppose $\{(y_\ell, \sigma_\ell, u_\ell)\}_\mathbb{N}$ is a sequence of global minima. The triple $(y_\ell, u_\ell, \sigma_\ell) = (0, -f_\ell, 0)$ is a feasible point for (4.2.4) and thus

$$J_\ell(y_\ell, u_\ell) \ \leq \ J_\ell(0, -f_\ell)\,.$$

As we consider nested finite elements subspaces, using the inverse triangle and Young's inequality we can deduce that the sequences $\{y_\ell\}_\mathbb{N}$ and $\{u_\ell\}_\mathbb{N}$ are bounded in $L^2(\Omega)$. If $\{(y_\ell, \sigma_\ell, u_\ell)\}_\mathbb{N}$ are only stationary points, the boundedness of $\{y_\ell\}_\mathbb{N}$ and $\{u_\ell\}_\mathbb{N}$ in $L^2(\Omega)$ is due to Assumption 5.1.3 (i).

Choosing $v_\ell = y_\ell$ in (4.4.6a) and $v_\ell = p_\ell$ in (4.4.6c) and taking into account (4.4.6b) and (4.4.7), as well as (2.1.1b), we find

$$\begin{aligned}
\gamma_a \|y_\ell\|^2_{1,\Omega} \leq a(y_\ell, y_\ell) &= (f + u_\ell, y_\ell)_{0,\Omega} \leq \Big(\|f\|_{0,\Omega} + \|u_\ell\|_{0,\Omega}\Big)\|y_\ell\|_{1,\Omega}\,,\\
\gamma_a \|p_\ell\|^2_{1,\Omega} \leq a(p_\ell, p_\ell) &= (y^d - y_\ell, p_\ell)_{0,\Omega} - \langle\langle \mu_\ell, p_\ell \rangle\rangle \leq\\
&\leq (y^d - y_\ell, p_\ell)_{0,\Omega} \leq \Big(\|y^d\|_{0,\Omega} + \|y_\ell\|_{0,\Omega}\Big)\|p_\ell\|_{1,\Omega}\,.
\end{aligned}$$

Due to the boundedness of $\{y_\ell\}_\mathbb{N}$ and $\{u_\ell\}_\mathbb{N}$ in $L^2(\Omega)$, the sequences $\{y_\ell\}_\mathbb{N}$ and $\{p_\ell\}_\mathbb{N}$ are bounded in V. Consequently, the sequences $\{\hat{\sigma}_\ell\}_\mathbb{N}$ and $\{\hat{\mu}_\ell\}_\mathbb{N}$ are bounded in V^*. Indeed, in view of (4.5.5),

$$\langle \hat{\sigma}_\ell, v\rangle \leq \|f + u_\ell\|_{0,\Omega}\|v\|_{0,\Omega} + \Gamma_a|y_\ell|_{1,\Omega}|v|_{1,\Omega} \leq (\|f + u_\ell\|_{0,\Omega} + \Gamma_a|y_\ell|_{1,\Omega})\,\|v\|_{1,\Omega},$$
$$\langle \hat{\mu}_\ell, v\rangle \leq \|y^d - y_\ell\|_{0,\Omega}\|v\|_{0,\Omega} + \Gamma_a|p_\ell|_{1,\Omega}|v|_{1,\Omega} \leq \Big(\|y^d - y_\ell\|_{0,\Omega} + \Gamma_a|p_\ell|_{1,\Omega}\Big)\,\|v\|_{1,\Omega},$$

and thus,

$$\|\hat{\sigma}_\ell\|_{V^*} \leq \|f + u_\ell\|_{0,\Omega} + \Gamma_a|y_\ell|_{1,\Omega},$$
$$\|\hat{\mu}_\ell\|_{V^*} \leq \|y^d - y_\ell\|_{0,\Omega} + \Gamma_a|p_\ell|_{1,\Omega}.$$

Since $L^2(\Omega)$, V, and V^*, are reflexive, separable Banach spaces, every bounded sequence admits a weakly convergent subsequence (see, e.g., [3]). Hence, there exist a subsequence $\mathbb{N}' \subseteq \mathbb{N}$ and a point $(y^*, \sigma^*, u^*, p^*, \mu^*) \in V \times V^* \times L^2(\Omega) \times V \times V^*$ such that for $\ell \in \mathbb{N}'$, $\ell \to \infty$ there holds

$$y_\ell \rightharpoonup y^* \text{ in } V, \qquad p_\ell \rightharpoonup p^* \text{ in } V, \tag{5.2.1a}$$
$$u_\ell \rightharpoonup u^* \text{ in } L^2(\Omega), \tag{5.2.1b}$$
$$\hat{\sigma}_\ell \rightharpoonup^* \sigma^* \text{ in } V^*, \quad \hat{\mu}_\ell \rightharpoonup^* \mu^* \text{ in } V^*. \tag{5.2.1c}$$

Since V is compactly embedded in $L^2(\Omega)$ (by the Rellich-Kondrachov Theorem, see, e.g., [1], p. 168), (5.2.1a) implies that

$$y_\ell \to y^* \text{ in } \ L^2(\Omega)\,, \quad p_\ell \to p^* \text{ in } L^2(\Omega)\,. \tag{5.2.2}$$

Restricting ourselves to a further subsequence, also denoted by $\mathbb{N}'$, for $\ell \to \infty$, $\ell \in \mathbb{N}'$ we get $y_\ell \to y^*$ and $p_\ell \to p^*$ pointwise almost everywhere. Hence, $y_\ell \leq 0$, $\ell \in \mathbb{N}'$ implies $y^* \leq 0$ a.e. in Ω.

Now, we show that the point $(y^*, \sigma^*, u^*, p^*, \mu^*)$ satisfies the state equations (3.3.11a), the adjoint state equation (3.3.11c), and equation (3.3.11d). Taking advantage of the assumption of limit density of the discrete subspaces, for any $v \in V$ we can find a sequence $v_\ell \in V_\ell$, $\ell \in \mathbb{N}$ such that $v_\ell \to v$ in V, $\ell \to \infty$. Due to (5.2.1), for $\ell \in \mathbb{N}'$, $\ell \to \infty$ we have

$$a(y_\ell, v_\ell) \to a(y^*, v), \qquad a(p_\ell, v_\ell) \to a(p^*, v),$$
$$(f + u_\ell, v_\ell)_{0,\Omega} \to (f + u^*, v)_{0,\Omega}, \qquad (y^d - y_\ell, v_\ell)_{0,\Omega} \to (y^d - y^*, v)_{0,\Omega},$$
$$\langle\langle \sigma_\ell, v_\ell\rangle\rangle = \langle \hat{\sigma}_\ell, v_\ell\rangle \to \langle \sigma^*, v\rangle, \qquad \langle\langle \mu_\ell, v_\ell\rangle\rangle = \langle \hat{\mu}_\ell, v_\ell\rangle \to \langle \mu^*, v\rangle.$$

Hence, passing to the limit in (4.4.6a) and (4.4.6c), we get

$$a(y^*, v) \;=\; (f + u^*, v)_{0,\Omega} - \langle \sigma^*, v\rangle,$$
$$a(p^*, v) \;=\; (y^d - y^*, v)_{0,\Omega} - \langle \mu^*, v\rangle.$$

Proposition 4.1.6 (*ii*) implies $u_\ell^d \to u^d$ in $L^2(\Omega)$, hence, in view of (4.4.6d) and (5.2.2), we deduce that for $\ell \in \mathbb{N}'$, $\ell \to \infty$

$$u_\ell \to u^* \text{ in } L^2(\Omega) \tag{5.2.3}$$

and that (p^*, u^*) satisfies (3.3.11d).

We further verify that $\sigma^* \in V_+^*$. According to Remark 5.1.2, $\{(V_\ell)_+\}_{\mathbb{N}}$ is limit dense in V_+. Hence, for any $v \in V_+$ there exists $v_\ell \in (V_\ell)_+$, $\ell \in \mathbb{N}$, such that $v_\ell \to v$ as $\ell \to \infty$. Due to $\sigma_\ell \in \mathcal{M}_+(\Omega)$ and (5.2.1c) it follows that $0 \le \langle\langle \sigma_\ell, v_\ell \rangle\rangle = \langle \hat{\sigma}_\ell, v_\ell \rangle \to \langle \sigma^*, v \rangle$, thus, $\langle \sigma^*, v \rangle \ge 0$ for any $v \in V_+$.

To show strong convergence of the sequence of the states in V, we observe that (4.2.2) implies

$$a(y_\ell, y_\ell) \le a(y_\ell, v_\ell) + (f + u_\ell, y_\ell - v_\ell)_{0,\Omega} \ , \ v_\ell \in K_\ell = (V_\ell)_-,$$

where $(V_\ell)_- := V_\ell \cap V_-$, $V_- := \{v \in V \,|\, v \le 0 \text{ a.e. in } \Omega\}$. Note that $\{(V_\ell)_-\}_{\mathbb{N}}$ is limit dense in V_-, and therefore, we can choose a sequence $\{v_\ell\}_{\mathbb{N}}$, $v_\ell \in (V_\ell)_-$, $\ell \in \mathbb{N}$ such that $v_\ell \to y^* \in V_-$ in V as $\ell \to \infty$. Now,

$$\begin{aligned}\gamma_a \|y_\ell - y^*\|_{1,\Omega}^2 &\le a(y_\ell - y^*, y_\ell - y^*) = a(y_\ell, y_\ell) - a(y_\ell, y^*) - a(y^*, y_\ell - y^*) \le \\ &\le a(y_\ell, v_\ell) + (f + u_\ell, y_\ell - v_\ell)_{0,\Omega} - a(y_\ell, y^*) - a(y^*, y_\ell - y^*).\end{aligned}$$

Due to the strong convergence of $\{y_\ell\}_{\mathbb{N}'}$ and $\{u_\ell\}_{\mathbb{N}'}$ in $L^2(\Omega)$, the expression on the right hand side converges to $a(y^*, y^*) + (f + u^*, y^* - y^*) - a(y^*, y^*) = 0$. Thus, for $\ell \in \mathbb{N}', \ell \to \infty$

$$y_\ell \to y^* \text{ in } V. \tag{5.2.4}$$

Consequently, due to (5.2.1c), (4.4.6b), and (4.4.6f), we get $0 = \langle \hat{\sigma}_\ell, y_\ell \rangle \longrightarrow \langle \sigma^*, y^* \rangle$ and $0 = \langle \hat{\mu}_\ell, y_\ell \rangle \longrightarrow \langle \mu^*, y^* \rangle$, and thus,

$$\langle \sigma^*, y^* \rangle = \langle \mu^*, y^* \rangle = 0.$$

Furthermore, due to the compact embedding of $L^2(\Omega)$ in V^*, (5.2.1b) implies that $\{u_\ell\}_{\mathbb{N}'}$ converges to u^* strongly in V^*. Since A is a bounded operator from V to V^*, (5.2.4) implies for $\ell \in \mathbb{N}', \ell \to \infty$

$$\begin{aligned}0 \le \|\hat{\sigma}_\ell - \sigma^*\|_{V^*} &\le \|Ay_\ell - Ay^*\|_{V^*} + \|u_\ell - u^*\|_{V^*} \le \\ &\le \|A\|_{\mathcal{L}(V,V^*)} \|y_\ell - y^*\|_V + \|u_\ell - u^*\|_{V^*} \to 0.\end{aligned}$$

and thus,

$$\hat{\sigma}_\ell \to \sigma^* \text{ in } V^*, \quad \ell \to \infty. \tag{5.2.5}$$

Combining (5.2.5) with (5.2.1a) and (4.4.6e), we get

$$\langle \sigma^*, p^* \rangle = 0.$$

We next show $\langle \mu^*, p^* \rangle \geq 0$. Setting $v_\ell = p_\ell$ in (4.4.6c) and observing that $\langle\langle \mu_\ell, p_\ell \rangle\rangle \geq 0$, we have

$$0 \geq a(p_\ell, p_\ell) - (y^d - y_\ell, p_\ell)_{0,\Omega}\,. \tag{5.2.6}$$

Since $I(v) := a(v, v)$ is a lower semicontinuous, convex functional on V, it is weakly lower semicontinuous and hence, in view of (5.2.1a) we get

$$a(p^*, p^*) \leq \liminf a(p_\ell, p_\ell)\,.$$

Moreover, due to (5.2.2) there holds

$$(y^d - y_\ell, p_\ell)_{0,\Omega} \to (y^d - y^*, p^*)_{0,\Omega}\,.$$

Hence, passing to the limit in (5.2.6) and taking into account that (y^*, u^*, σ^*) satisfies (3.2.1b), it follows that

$$0 \geq a(p^*, p^*) - (y^d - y^*, p^*)_{0,\Omega} = -\langle \mu^*, p^* \rangle\,,$$

whence $\langle \mu^*, p^* \rangle \geq 0$.

Let us prove that $\langle \sigma^*, (p^*)^+ \rangle = \langle \sigma^*, (p^*)^- \rangle = 0$. As p_ℓ, $p^* \in V$, $\ell \in \mathbb{N}'$, there holds $(p_\ell)^+, (p_\ell)^-, (p^*)^+, (p^*)^- \in V$. Thus, (5.2.1a) yields $(p_\ell)^+ \rightharpoonup (p^*)^+ \in V$, $(p_\ell)^- \rightharpoonup (p^*)^- \in V$ in V as $\ell \to \infty$ (see, e.g., [45], p. 159). Together with (4.4.6e), this leads to

$$0 = \langle\langle \sigma_\ell, (p_\ell)^+ \rangle\rangle \to \langle \sigma^*, (p^*)^+ \rangle, \quad 0 = \langle\langle \sigma_\ell, (p_\ell)^- \rangle\rangle \to \langle \sigma^*, (p^*)^- \rangle.$$

Consequently, $\langle \sigma^*, (p^*)^+ \rangle = \langle \sigma^*, (p^*)^- \rangle = 0$, and, by Corollary 2.4.8, $p^* = 0$ a.e. in $\mathcal{C}^*$, $\mathcal{C}^* = \operatorname{int}(\operatorname{supp}(\sigma^*))$.

Finally, we complete the proof by verifying conditions (5.1.1) and (3.3.13).

To prove (5.1.1), let $v \in C^1(\overline{\Omega}) \subset H^1(\Omega)$ be arbitrary but fixed. We then have $y^* v \in V$ (compare with [29], p. 21). Due to the limit density of $\{S_\ell\}_{\mathbb{N}}$ in $H^1(\Omega)$, there exists a sequence $\{v_\ell\}_{\mathbb{N}}$, $v_\ell \in S_\ell$, $\ell \in \mathbb{N}$ such that $v_\ell \to v$ in $H^1(\Omega)$ and thus in V. As $v_\ell \in C(\overline{\Omega})$ and $y_\ell \in C_0(\Omega)$, we have $v_\ell\, y_\ell \in C_0(\Omega)$. Moreover, $(v_\ell\, y_\ell)|_T \in H_1(T)$ for all $T \in \mathcal{T}_\ell$, therefore, $v_\ell\, y_\ell \in V$, $\ell \in \mathbb{N}$. With (5.2.4), we then have

$$y_\ell\, v_\ell \to y^* v \text{ in } V, \quad \ell \to \infty.$$

For all $\ell \in \mathbb{N}$, we have $(y_\ell\, v_\ell)(a) = y_\ell(a) v_\ell(a)$, $a \in \mathcal{N}_\ell$, thus, $(y_\ell\, v_\ell)(a) = 0$, $a \in A_\ell$. Therefore, in view of (5.2.1c),

$$0 = \langle \hat{\mu}_\ell, y_\ell\, v_\ell \rangle \to \langle \mu^*, y^* v \rangle, \quad \ell \to \infty,$$

hence, $\langle \mu^*, y^* v \rangle = 0$. As v was chosen arbitrary in $C^1(\overline{\Omega})$, we have

$$\langle \mu^*, y^* v \rangle = 0, \quad \forall\, v \in C^1(\overline{\Omega}).$$

In order to prove (3.3.13), we will use the following observations. On one hand, due to (4.4.6f), $\forall\, \ell \in \mathbb{N}$ we have

$$\langle \hat{\mu}_\ell, v_\ell \rangle = 0, \quad \forall\, v_\ell \in V_\ell \cap V_{\mathcal{I}_\ell \cup \mathcal{F}_\ell(y_\ell)}. \tag{5.2.7}$$

On the other hand, we know that $\{y_\ell\}_{\mathbb{N}'}$ converges to y^* pointwise a.e. in Ω. Therefore, for sufficiently large $\bar{\ell}_1$ there holds

$$y_\ell < 0 \text{ a.e. in } \mathcal{I}^*, \quad \ell \geq \bar{\ell}_1,$$

i.e., $\mathcal{I}^* \subseteq \mathcal{I}_\ell$, $\ell \geq \bar{\ell}_1$. For $\ell \geq \bar{\ell}_1$, construct the set

$$\widetilde{\mathcal{I}}_\ell := \bigcup \{T \in \mathcal{T}_\ell \mid \operatorname{int}(T) \subseteq \mathcal{I}^*\},$$

which evidently satisfies $\widetilde{\mathcal{I}}_\ell \subseteq \mathcal{I}^* \subseteq \mathcal{I}_\ell$. Note that $\widetilde{\mathcal{I}}_\ell$ can be empty. In order to avoid this trivial situation, choose $\bar{\ell}_2$ sufficiently large so that $\widetilde{\mathcal{I}}_\ell \neq \varnothing$ for $\ell \geq \bar{\ell}_2$. Setting $\bar{\ell} := \max\{\bar{\ell}_1, \bar{\ell}_2\}$, we thus have selected clusters of triangles $\widetilde{\mathcal{I}}_\ell$, $\ell \geq \bar{\ell}$ that satisfy

$$\varnothing \neq \widetilde{\mathcal{I}}_\ell \subseteq \mathcal{I}^* \subseteq \mathcal{I}_\ell\,, \quad \ell \geq \bar{\ell}. \tag{5.2.8}$$

Figure 5.1 shows an example of the above construction with $\bar{\ell}_1 = \ell$ and $\bar{\ell}_2 = \ell + 1$.

Let $v \in C_{\mathcal{I}^*,0} := \{v \in C_0(\Omega) \mid v|_{\mathcal{I}^*} \in C_0^\infty(\mathcal{I}^*), v|_{\Omega \setminus \mathcal{I}^*} = 0\}$ be arbitrary but fixed. There holds $\operatorname{supp}(v) \subseteq \mathcal{I}^*$. Since $\operatorname{supp}(v)$ is closed and $\mathcal{I}^*$ is open, there is a gap between $\partial\mathcal{I}^*$ and $\partial \operatorname{supp}(v)$, therefore, we can find $\ell(v) \in \mathbb{N}'$, $\ell(v) \geq \bar{\ell}$ such that the cluster of triangles $\widetilde{\mathcal{I}}_{\ell(v)} \subseteq \mathcal{I}^*$ will contain the support of v, i.e.,

$$\operatorname{supp}(v) \subseteq \widetilde{\mathcal{I}}_\ell \subseteq \mathcal{I}^* \subseteq \mathcal{I}_\ell\,, \quad \ell \geq \ell(v).$$

(In the example in Figure 5.1 (c), $\ell(v) = \ell + 2$.) Evidently, $v \in V_{\widetilde{\mathcal{I}}_{\ell(v)},0} \subseteq V_{\mathcal{I}^*,0}$. Moreover, there obviously holds $\widetilde{\mathcal{I}}_{\ell(v)} \subseteq \mathcal{I}_\ell$, $\ell \geq \ell(v)$, and, thus,

$$V_\ell \cap V_{\widetilde{\mathcal{I}}_{\ell(v)}} \subseteq V_\ell \cap V_{\mathcal{I}_\ell \cup \mathcal{F}_\ell(y_\ell)}, \quad \ell \geq \ell(v).$$

Hence, due to (5.2.7),

$$\langle \hat{\mu}_\ell, v_\ell \rangle = 0, \quad \forall\, v_\ell \in V_\ell \cap V_{\widetilde{\mathcal{I}}_{\ell(v)}}, \quad \ell \geq \ell(v). \tag{5.2.9}$$

The sequence of subspaces $\{V_\ell \cap V_{\widetilde{\mathcal{I}}_{\ell(v)}}\}_{\ell \geq \ell(v)} \subset V_{\widetilde{\mathcal{I}}_{\ell(v)}}$ is limit dense in $V_{\widetilde{\mathcal{I}}_{\ell(v)}}$. Therefore, there exists a sequence $\{v_\ell\}_{\ell \geq \ell(v)}$, $v_\ell \in V_\ell \cap V_{\widetilde{\mathcal{I}}_{\ell(v)}}$ such that $v_\ell \to v$ in V. Hence, by (5.2.9) and (5.2.1c), for $\ell \in \mathbb{N}'$, $\ell \geq \ell(v)$, $\ell \to \infty$ we have

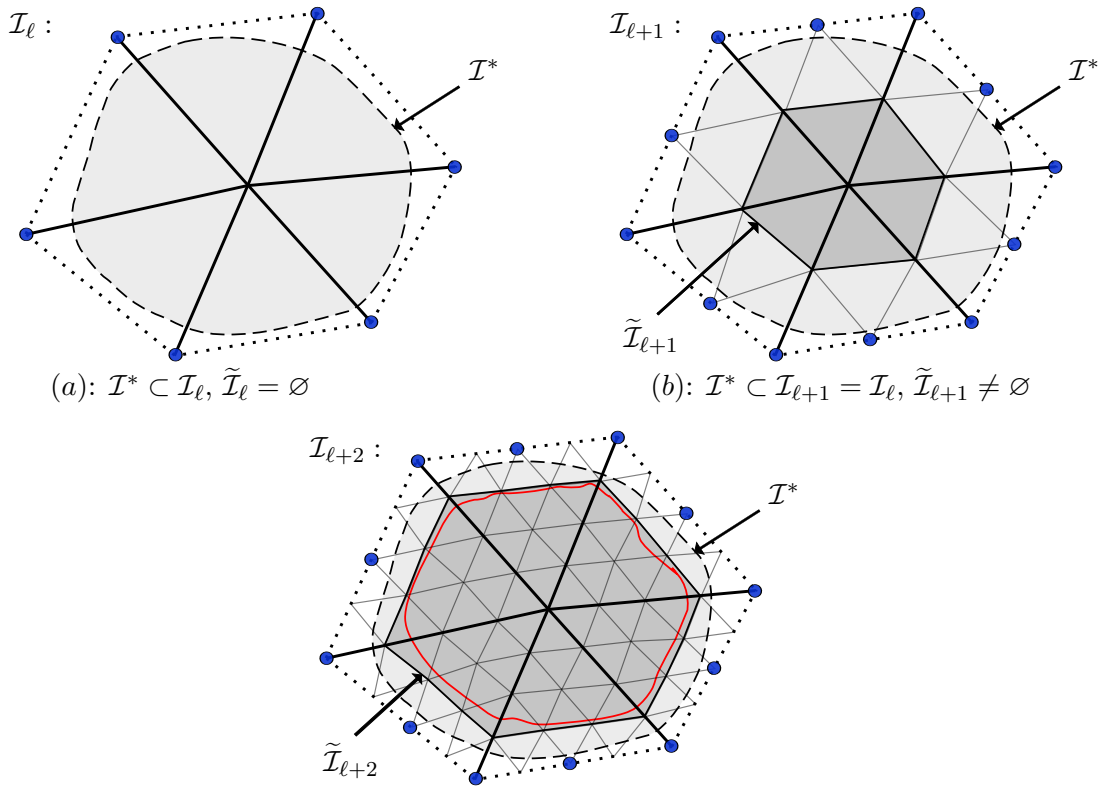
$$0 = \langle \hat{\mu}_\ell, v_\ell \rangle \to \langle \mu^*, v \rangle,$$

and so $\langle \mu^*, v \rangle = 0$. As v can be chosen arbitrary in $C_{\mathcal{I}^*,0}$, we conclude that

$$\langle \mu^*, v \rangle = 0, \ \forall\, v \in C_{\mathcal{I}^*,0}.$$

Finally, the density of $C_{\mathcal{I}^*,0}$ in $V_{\mathcal{I}^*,0}$ implies $\langle \mu^*, v \rangle = 0$, $\forall\, v \in V_{\mathcal{I}^*,0}$.

The proof of the theorem is complete. □



(a): $\mathcal{I}^* \subset \mathcal{I}_\ell$, $\widetilde{\mathcal{I}}_\ell = \varnothing$

(b): $\mathcal{I}^* \subset \mathcal{I}_{\ell+1} = \mathcal{I}_\ell$, $\widetilde{\mathcal{I}}_{\ell+1} \neq \varnothing$

(c): $\mathcal{I}^* \subset \mathcal{I}_{\ell+2} = \mathcal{I}_{\ell+1}$, $\mathrm{supp}(v) \subset \widetilde{\mathcal{I}}_{\ell+2}$

Figure 5.1: Figures (a) and (b) show an example with $\bar{\ell}_1 = \ell$, $\bar{\ell}_2 = \ell + 1$. Blue points indicate the active nodes of the mesh. The dotted line represents $\partial\mathcal{I}_\ell = \partial\mathcal{I}_{\ell+1}$. The dashed line shows $\partial\mathcal{I}^*$. On the mesh ℓ (picture (a)) the inclusion $\mathcal{I}^* \subset \mathcal{I}_\ell$ holds (i.e., $\bar{\ell}_1 = \ell$), whereas $\widetilde{\mathcal{I}}_\ell = \varnothing$. One regular refinement results in (b), which is a realization of (5.2.8), i.e., $\bar{\ell} = \bar{\ell}_2 = \ell + 1$. Another refinement leads to the configuration depicted in Figure (c). The red line represents $\partial(\mathrm{supp}(v))$ of a function $v \in C_0^\infty(\Omega)$ which satisfies $\ell(v) = \ell + 2$.

Remark 5.2.1. *(Convergence to C-stationary points of other kinds)*
The concepts of ε-almost, almost, and C-stationarity differ only by the local characterization of the multiplier $\mu^* \in V^*$. Any almost C-stationary point is an ε-almost C-stationary point. If, additionally to the assumptions of Theorem 5.1.4, we require $\mathcal{I}^*$ to be Lipschitz, we would have $V_{\mathcal{I}^*,0} = V_{\mathcal{I}^*}$ (see Remark 2.3.8) and the limit point would be a C-stationary point. For arbitrary $\mathcal{I}^*$, the construction suggested in the final part of the proof of Theorem 5.1.4 cannot be used to prove that the limit point is a C-stationary point as, in general, $C_{\mathcal{I}^*,0}$ is not dense in $V_{\mathcal{I}^*}$.

Remark 5.2.2. It would be desirable to extend the result of Theorem 5.1.4 by showing

(i) $p_\ell \to p^*$ in V;

(*ii*) $\mu_\ell \to \mu^*$ in V^*;

(*iii*) $\langle \mu^*, p^* v \rangle \geq 0$, $\forall\, v \in C^1(\overline{\Omega}) \cap V_+$ (cf. (4.4.15) and Remark 4.4.13);

(*iv*) convergence to an almost S-stationary point or some other infinite-dimensional analog of the discrete S-stationary point, assuming sequences of discrete S-stationary points.

Note that (*ii*) and (*iii*) can be deduced from (*i*) just as the strong convergence in V^* for $\hat{\sigma}_\ell$ and (5.1.1) are deduced from the strong convergence of the sequence of the discrete states in the proof of Theorem 5.1.4. The pointwise a.e. convergence of p_ℓ to p^* and the fact that, due to $\hat{\mu}_\ell \rightharpoonup^* \mu$ in V^*, we have $\|\mu\|_{V^*} \leq \liminf \|\hat{\mu}_\ell\|_{V^*}$ (see [55], p. 56), could possibly support for the proof of (*iv*). Note that, in order to show convergence to a S-stationarity point, we need to choose an appropriate definition of non-negativity of p and μ in $\mathcal{B}$.

5.3 Discussion of the convergence result

In this paragraph, the result of Theorem 5.1.4 will be discussed in view of the non-uniqueness of the global solutions of the optimal control problem (a fortiory stationary points of any kind).

For the sake of convenience, we introduce the notation $U_\ell := (y_\ell, \sigma_\ell, u_\ell, p_\ell, \mu_\ell)$, $U := (y, \sigma, u, p, \mu)$. In general, U_ℓ and U can signify any kind of stationary point of the discrete and continuous optimal control problems. Further, let us denote the sets of global minimizers, local minimizers, and C-stationary points of the discrete optimal control problem (4.2.2) by SOL_ℓ^{glob}, SOL_ℓ^{loc}, and SOL_ℓ^{C}. On each mesh level $\ell \in \mathbb{N}$, (4.2.2) admits at least one global minimizer (cf. Theorem 4.2.1), thus, in view of Corollary 4.4.6 there obviously holds

$$\mathrm{SOL}_\ell^{glob} \subseteq \mathrm{SOL}_\ell^{loc} \subseteq \mathrm{SOL}_\ell^{C}\,.$$

Similarly, we denote the sets of global minimizers, local minimizers, and almost C-stationary points of the continuous optimal control problem (3.1.1) by SOL^{glob}, SOL^{loc}, and SOL^{C}, respectively. These, in turn, satisfy

$$\mathrm{SOL}^{glob} \subseteq \mathrm{SOL}^{loc} \subseteq \mathrm{SOL}^{C}\,.$$

In this notation, Theorem 5.1.4 reads as follows:

For a given sequence $\{U_\ell\}_{\mathbb{N}}$, $U_\ell \in \mathrm{SOL}_\ell^C$, $\ell \in \mathbb{N}$ there exist a subsequence $\mathbb{N}' \subseteq \mathbb{N}$ and a point $U \in \mathrm{SOL}^C$, such that $\{U_\ell\}_{\mathbb{N}'}$ converges to U in the sense of Theorem 5.1.4.

For the sake of simplicity, let us assume that in the continuous setting and on each mesh level $\ell \in \mathbb{N}$ of the discrete setting there are countably many almost C-stationary and discrete C-stationary points, respectively, i.e., $\mathrm{SOL}^C = \{U^{(i)}\}_{i \in \mathbb{N}}$ and

$\mathrm{SOL}_\ell^C = \{U_\ell^{(i)}\}_{i\in\mathbb{N}}$, $\ell \in \mathbb{N}$. We further assume that for any $\ell \in \mathbb{N}$, the kth discrete C-stationarity point $(U_\ell^{(k)})$ provides the closest approximation to the almost C-stationary point $U^{(k)}$. Every element of SOL_ℓ^C satisfies the discrete C-stationarity system on the level $\ell \in \mathbb{N}$, just as every element of SOL^C the continuous almost C-stationarity system. As we do not have any means to distinguish between different C-stationarity points on a fixed mesh level and different almost C-stationary points in the continuous setting, the sequence $\{U_\ell\}_\mathbb{N}$ addressed in Theorem 5.1.4 can be composed of discrete C-stationarity points approximating different solutions of the continuous system. For instance, $\{U_\ell\}_\mathbb{N} = \{U_1^{(7)}, U_2^{(2)}, U_3^{(1)}, \ldots\}$. If the assumptions of Theorem 5.1.4 are satisfied, a subsequence $\mathbb{N}_1'$ of this sequence shall converge to a unique limit point, say $U^{(1)} \in \mathrm{SOL}^C$. In view of the latter observations, we can give a more accurate variant of Theorem 5.1.4:

For a given sequence $\{U_\ell\}_\mathbb{N}$, $U_\ell \in \mathrm{SOL}_\ell^C$, $\ell \in \mathbb{N}$ there exist a subsequence $\mathbb{N}_1' \subseteq \mathbb{N}$ and a point $U^{(1)} \in \mathrm{SOL}^C$, such that $\{U_\ell\}_{\mathbb{N}_1'}$ converges to $U^{(1)}$ in the sense of Theorem 5.1.4. In general, there is no uniqueness, i.e., there can be further subsequences $\mathbb{N}_i' \subseteq \mathbb{N}$, $i \geq 1$ with limit points in the sense of Theorem 5.1.4, $U^{(i)} \in \mathrm{SOL}^C$, such that $U^{(i)} \neq U^{(1)}$, $i \geq 1$.

Note that in the worst case scenario none of the limit points delivers a global or local minimum to the optimal control problem, i.e., $U^{(i)} \notin \mathrm{SOL}^{glob}$, $U^{(i)} \notin \mathrm{SOL}^{loc}$, $\forall\, i$.

Remark 5.3.1. The discrete and the continuous optimal control problems always have S-stationary points. Thus, in practice, we are free to consider sequences of discrete S-stationary points. For such sequences, Theorem 5.1.4 guarantees convergence to an almost C-stationary point but does not say whether the limit point is also an S-stationary point of the continuous problem.

6 A posteriori error analysis

This chapter is devoted to the derivation of a residual type error estimate for the difference between the solutions of the continuous and the discrete optimal control problem and to an analysis of the components of the derived estimate.

6.1 Residual-type a posteriori error estimate

In the a posteriori error analysis, one typically aims to prove two-sided estimates in the form

$$\text{“estimator” + [...] } \lesssim \text{ “error” } \lesssim \text{ “estimator” + [...] ,}$$

where “error” stands for a certain measure of the difference between the unknown exact solution and the approximate solution, “estimator” stands for a fully computable expression that involves the approximate solution and the data of the problem, and the square brackets may contain terms that should be essentially dominated by the estimator. The inequality on the right-hand side of the estimate is called *the reliability estimate*, the inequality on the left-hand side - *the efficiency estimate*. These names emphasize the qualities that the estimator acquires in view of the corresponding estimate. Indeed, the estimate on the right-hand side implies that if the estimator is smaller then a certain tolerance, than so is the error. Thus, the estimator is reliable. On the other hand, the estimate on the left-hand side guarantees that the estimator cannot overestimate the error by a too big factor, i.e., the estimator is efficient. From the perspective of solving the underlying problem on sequences of nested meshes, the efficiency estimate additionally assures that the error will not decrease faster than the estimator.

In the following, we will derive a residual-type two-sided a posteriori error estimate for the discretization errors in the state, the adjoint state and the control

$$e_{\ell,y} := y - y_\ell, \quad e_{\ell,p} := p - p_\ell, \quad e_{\ell,u} := u - u_\ell,$$

in terms of the total discretization error $e_\ell := (e_{\ell,y}, e_{\ell,p}, e_{\ell,u})$ measured in the norm

$$|||e_\ell|||^2 := \|e_{\ell,y}\|^2_{1,\Omega} + \|e_{\ell,p}\|^2_{1,\Omega} + \|e_{\ell,u}\|^2_{0,\Omega}.$$

In particular, it will be shown that

$$\eta_\ell^2 - osc_{\ell,eff}^2 - e_{\ell,eff}^c \lesssim |||e_\ell|||^2 \lesssim \eta_\ell^2 + osc_{\ell,rel}^2 + e_{\ell,rel}^c\,, \tag{6.1.1}$$

where η_ℓ is the residual-type a posteriori error estimator, whereas $e^c_{\ell,rel}$, $e^c_{\ell,eff}$ and $osc_{\ell,rel}$, $osc_{\ell,eff}$ stand for the consistency errors and data oscillations associated with the reliability and efficiency estimates, respectively.

Remark 6.1.1. The proof of the error estimate (6.1.1) will rely on a set of properties fulfilled by all types of discrete and continuous stationary points introduced in the earlier chapters. More precisely, we assume that the point $(y_\ell, \sigma_\ell, u_\ell, p_\ell, \mu_\ell) \in V_\ell \times \mathcal{M}_\ell \times S_\ell \times V_\ell \times \mathcal{M}_\ell$ satisfies the system

$$a(y_\ell, v_\ell) = (f + u_\ell, v_\ell)_{0,\Omega} - \langle \tilde{\sigma}_\ell, v_\ell \rangle \,, \ \forall\, v_\ell \in V_\ell, \tag{6.1.2a}$$
$$\psi_\ell - y_\ell \in V_+, \ \sigma_\ell \in \mathcal{M}_\ell \cap \mathcal{M}_+(\overline{\Omega}), \ \langle \sigma_\ell, \psi_\ell - y_\ell \rangle = 0, \tag{6.1.2b}$$
$$a(p_\ell, v_\ell) = (y^d - y_\ell, v_\ell)_{0,\Omega} - \langle \tilde{\mu}_\ell, v_\ell \rangle \,, \ \forall\, v_\ell \in V_\ell, \tag{6.1.2c}$$
$$p_\ell = \alpha(u_\ell - u^d_\ell), \tag{6.1.2d}$$
$$p_\ell(a) = 0, \ a \in C_\ell, \tag{6.1.2e}$$
$$\mu_\ell(a) = 0, \ a \in I_\ell \tag{6.1.2f}$$

and $(y, \sigma, u, p, \mu) \in V \times V^* \times L^2(\Omega) \times V \times V^*$ satisfies

$$a(y, v) = (f + u, v)_{0,\Omega} - \langle \sigma, v \rangle \,, \ \forall\, v \in V, \tag{6.1.3a}$$
$$\psi - y \in V_+, \ \sigma \in V^*_+, \ \langle \sigma, \psi - y \rangle = 0, \tag{6.1.3b}$$
$$a(p, v) = (y^d - y, v)_{0,\Omega} - \langle \mu, v \rangle \,, \ \forall\, v \in V, \tag{6.1.3c}$$
$$p = \alpha(u - u^d), \tag{6.1.3d}$$
$$p = 0 \text{ a.e. in } \mathcal{C}, \tag{6.1.3e}$$
$$\text{“}\mu = 0 \text{ in } \mathcal{I}\text{”}. \tag{6.1.3f}$$

The quotation marks in (6.1.3f) emphasize that, though the localization of the Lagrangian multiplier μ is important for the error estimator, the latter does not make an explicit use of the fine differences between the $\varepsilon-$almost, almost-, and straight stationarity concepts. Furthermore, the aspect of the classification of stationary points that specifies how the adjoint state and the multiplier associated with the adjoint state equation interact in the biactive set is absent in both systems. Hence, the a posteriori error estimate (6.1.1) is applicable to all types of the discrete and the continuous stationary points.

6.2 Components of the reliability and efficiency estimates

In this section, we will introduce the consistency errors and data oscillations associated with the reliability and efficiency estimates, as well as the residual-type a posteriori error estimator consisting of the element and edge residuals.

6.2.1 Residual-type a posteriori error estimator

The residual-type a posteriori error estimator is given by

$$\eta_\ell \;:=\; \eta_\ell(y) + \eta_\ell(p), \tag{6.2.1}$$

where $\eta_\ell(y)$ and $\eta_\ell(p)$ consist of element residuals and edge residuals associated with the state equation (6.1.2a) and the adjoint state equation (6.1.2a)

$$\eta_\ell(y) := \Big(\sum_{T\subseteq\mathcal{Z}_\ell} \eta_T^2(y) + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} \eta_E^2(y) \Big)^{1/2}, \tag{6.2.2a}$$

$$\eta_\ell(p) := \Big(\sum_{T\subseteq\mathcal{I}_\ell} \eta_T^2(p) + \sum_{E\in\mathcal{E}_{\mathcal{I}_\ell}} \eta_E^2(p) \Big)^{1/2}. \tag{6.2.2b}$$

In particular, the element residuals $\eta_T(y)$, $\eta_T(p)$ are weighted element-wise L^2-residuals with respect to the strong form of the discrete state and discrete adjoint state equations

$$\eta_T(y) := h_T \| f + u_\ell \|_{0,T}, \tag{6.2.3a}$$

$$\eta_T(p) := h_T \| y^d - y_\ell \|_{0,T}, \tag{6.2.3b}$$

whereas the edge residuals $\eta_E(y)$, $\eta_E(p)$ are weighted L^2-norms of the jumps $\nu_E \cdot [\nabla y_\ell]$ and $\nu_E \cdot [\nabla p_\ell]$ of the normal derivatives across the interior edges

$$\eta_E(y) := h_E^{1/2} \| \nu_E \cdot [\nabla y_\ell] \|_{0,E}, \tag{6.2.4a}$$

$$\eta_E(p) := h_E^{1/2} \| \nu_E \cdot [\nabla p_\ell] \|_{0,E}. \tag{6.2.4b}$$

6.2.2 Data oscillations

We distinguish between reliability and efficiency related data oscillations. The data oscillation associated with the reliability estimate is given by

$$osc_{\ell,rel}^2 := \sum_{T\subseteq\mathcal{T}_\ell} osc_T^2(u^d), \quad \text{where} \;\; osc_T(u^d) := \| u^d - u_\ell^d \|_{0,T}, \tag{6.2.5}$$

whereas the data oscillation associated with the efficiency estimate reads

$$osc_{\ell,eff}^2 := osc_{\mathcal{Z}_\ell}^2(f) + osc_{\mathcal{I}_\ell}^2(y^d), \tag{6.2.6}$$

where

$$osc_{\mathcal{Z}_\ell}^2(f) := \sum_{T\subseteq\mathcal{Z}_\ell} osc_T^2(f), \quad osc_T(f) := h_T \| f - f_\ell \|_{0,T}, \tag{6.2.7a}$$

$$osc_{\mathcal{I}_\ell}^2(y^d) := \sum_{T\subseteq\mathcal{I}_\ell} osc_T^2(y^d), \quad osc_T(y^d) := h_T \| y^d - y_\ell^d \|_{0,T}. \tag{6.2.7b}$$

6.2.3 Consistency errors

Let us denote the discretization error in the Lagrangian multipliers by

$$e_{\ell,\sigma} \in V^*, \quad e_{\ell,\sigma}(v) := \langle \tilde{\sigma}_\ell - \sigma, v \rangle, \ \forall v \in V,$$
$$e_{\ell,\mu} \in V^*, \quad e_{\ell,\mu}(v) := \langle \tilde{\mu}_\ell - \mu, v \rangle, \ \forall v \in V.$$

As in case of the data oscillation, we distinguish between reliability and efficiency related consistency errors.

Consistency error associated with the reliability estimate

The error analysis involves error terms reflecting the mismatch between the continuous and discrete active and strongly active sets and the approximation of y by y_ℓ and p by p_ℓ, namely,

$$e^{(1)}_{\ell,\sigma} := e_{\ell,\sigma}(e_{\ell,y}), \quad e^{(2)}_{\ell,\sigma} := -e_{\ell,\sigma}(e_{\ell,p}), \tag{6.2.8a}$$
$$e^{(1)}_{\ell,\mu} := e_{\ell,\mu}(e_{\ell,y}), \quad e^{(2)}_{\ell,\mu} := e_{\ell,\mu}(e_{\ell,p}). \tag{6.2.8b}$$

These quantities constitute to *the consistency error related to the reliability estimate* (or the primal-dual mismatch in complementarity):

$$e^c_{\ell,rel} := e^{(1)}_{\ell,\sigma} + e^{(2)}_{\ell,\sigma} + e^{(1)}_{\ell,\mu} + e^{(2)}_{\ell,\mu}. \tag{6.2.9}$$

Consistency error associated with the efficiency estimate

The consistency error related to the efficiency estimate, $e^c_{\ell,eff}$, measures the mismatch between the continuous and the discrete active and strongly active sets, but, unlike the consistency error in the reliability, consists of contributions involving the exact multipliers σ and μ, localized to the areas where the discrete and the continuous sets are not matching. The expression of $e^c_{\ell,eff}$ involves the so-called *bubble functions* (see, e.g., [61]).

Definition 6.2.1. *(Bubble functions)*
Denote by λ_1^T, λ_2^T, λ_3^T the barycentric coordinates of $T \in \mathcal{T}_\ell$. The element bubble functions are then defined by

$$b_T := \begin{cases} 27 \prod_{i=1}^3 \lambda_i^T, & \text{on } T \\ 0 \qquad\qquad\quad , & \text{on } \Omega \setminus T \end{cases}.$$

If in the triangles T_+ and T_-, $T_+ \cup T_- = \omega_E$, $E \in \mathcal{E}_\ell$, the vertices are enumerated so that $\mathcal{N}_\ell(E)$ are numbered first, then the edge bubble functions are given by

$$b_E := \begin{cases} 4 \prod_{i=1}^2 \lambda_i^{T_\pm}, & \text{on } T_\pm \\ 0 \qquad\qquad\quad , & \text{on } \Omega \setminus \omega_E \end{cases}.$$

Remark 6.2.2. Clearly, the element and the edge bubble functions satisfy $b_T|_T \geq 0$, $\mathrm{supp}(b_T) = T$, $b_T|_{\partial T} = 0$ and $b_E|_{\omega_E} \geq 0$, $\mathrm{supp}(b_E) = \omega_E$, $b_E|_{\partial\omega_E} = 0$, i.e., $b_T \in V_{T,0} \cap V_+$ and $b_E \in V_{\omega_E,0} \cap V_+$.

The consistency error associated with the efficiency error estimate is given as follows

$$e^c_{\ell,eff} := \left[\sum_{T\subseteq\mathcal{Z}_\ell} e^\sigma_T \; + \sum_{T\subseteq\mathcal{I}_\ell} e^\mu_T \; + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} e^\sigma_{\omega_E} \; + \sum_{E\in\mathcal{E}_{\mathcal{I}_\ell}} e^\mu_{\omega_E} \right]^2, \tag{6.2.10}$$

where

$$e^\sigma_T := \frac{\langle \sigma, \psi^\sigma_T\rangle}{|\psi^\sigma_T|_{1,T}}, \;\; e^\mu_T := \frac{\langle \mu, \psi^\mu_T\rangle}{|\psi^\mu_T|_{1,T}}, \;\; e^\sigma_{\omega_E} := -\frac{\langle \sigma, \psi^\sigma_E\rangle}{|\psi^\sigma_E|_{1,\omega_E}}, \;\; e^\mu_{\omega_E} := -\frac{\langle \mu, \psi^\mu_E\rangle}{|\psi^\mu_E|_{1,\omega_E}} \tag{6.2.11}$$

with ψ^σ_T, $\psi^\mu_T \in C(\overline{\Omega}) \cap V_{T,0}$ and ψ^σ_E, $\psi^\mu_E \in C(\overline{\Omega}) \cup V_{\omega_E,0}$ defined according to

$$\begin{aligned} \psi^\sigma_T &:= (f_\ell + u_\ell) b_T, \quad \psi^\sigma_E := \nu_E \cdot [\nabla y_\ell] b_E, \\ \psi^\mu_T &:= (y^d_\ell - y_\ell) b_T, \quad \psi^\mu_E := \nu_E \cdot [\nabla p_\ell] b_E. \end{aligned}$$

6.3 Proof of the reliability estimate

Theorem 6.3.1. *Let (y, σ, u, p, μ) and $(y_\ell, \sigma_\ell, u_\ell, p_\ell, \mu_\ell)$ be solutions of (6.1.3) and (6.1.2), respectively, and let η_ℓ, $osc_{\ell,rel}$, and $e^c_{\ell,rel}$ be the error estimator, the data oscillation, and the consistency error as given by (6.2.1), (6.2.5) and (6.2.9). Then, it holds*

$$|||e_\ell|||^2 \lesssim \eta^2_\ell + osc^2_{\ell,rel} + e^c_{\ell,rel}\,. \tag{6.3.1}$$

The proof of Theorem 6.3.1 will be given by a series of lemmata.

We note that due to the presence of u, u_ℓ in the right-hand sides of the continuous and the discrete state equations (6.1.2a), (6.1.3a) and of y, y_ℓ in the right-hand sides of the continuous and the discrete adjoint state equations (6.1.2c), (6.1.3c), neither y_ℓ nor p_ℓ satisfy Galerkin orthogonality. As in the case of the a posteriori error analysis of finite element approximations of control and/or state constrained distributed optimal control problems for second order elliptic PDEs (cf., e.g., [43]), Galerkin orthogonality can be achieved with respect to an auxiliary state and an auxiliary adjoint state.

Definition 6.3.2. *(Auxiliary state and auxiliary adjoint state)*
The auxiliary state $y(u_\ell) \in V$ and the auxiliary adjoint state $p(y_\ell) \in V$ are the unique exact solutions of the infinite dimensional problems

$$a(y(u_\ell), v) = (f + u_\ell, v)_{0,\Omega} - \langle \tilde{\sigma}_\ell, v\rangle, \;\; v \in V, \tag{6.3.2a}$$

$$a(p(y_\ell), v) = (y^d - y_\ell, v)_{0,\Omega} - \langle \tilde{\mu}_\ell, v\rangle, \;\; v \in V, \tag{6.3.2b}$$

respectively.

Indeed, from (6.3.2a), (6.1.2a) and (6.3.2b), (6.1.2c) it follows easily that

$$a(y(u_\ell) - y_\ell, v_\ell) = 0,\ v_\ell \in V_\ell, \tag{6.3.3a}$$
$$a(p(y_\ell) - p_\ell, v_\ell) = 0,\ v_\ell \in V_\ell. \tag{6.3.3b}$$

Lemma 6.3.3. *Let the assumptions of Theorem 6.3.1 be satisfied and let $y(u_\ell)$, $p(y_\ell)$ be the auxiliary state and the auxiliary adjoint state as given by* (6.3.2)*. Further, let $\eta_\ell(y)$ and $\eta_\ell(p)$ be the components of the error estimate given by* (6.2.2)*. Then there holds*

$$\|y(u_\ell) - y_\ell\|_{1,\Omega} \lesssim \eta_\ell(y), \tag{6.3.4a}$$
$$\|p(y_\ell) - p_\ell\|_{1,\Omega} \lesssim \eta_\ell(p). \tag{6.3.4b}$$

Proof. Set $e := y(u_\ell) - y_\ell$. There holds

$$a(e, e) = r(e - P_{\ell,C}e), \tag{6.3.5}$$

where $P_{\ell,C}$ is Clément's quasi-interpolation operator (see Definition 4.1.3) and $r(\cdot)$ is the residual of the auxiliary state equation (6.3.2a) with respect to the approximation y_ℓ:

$$r(v) \ = \ (f + u_\ell, v)_{0,\Omega} - \langle \tilde{\sigma}_\ell, v\rangle - a(y_\ell, v)\ ,\ v \in V.$$

It is easy to verify (6.3.5). Indeed, according to Proposition 4.5.4, $\langle \tilde{\sigma}_\ell, v_\ell\rangle = \langle \hat{\sigma}_\ell, v_\ell\rangle$ for all $v_\ell \in V_\ell$. As a consequence, $r(v_\ell) = 0$ for all $v_\ell \in V_\ell$, i.e., y_ℓ is a Galerkin approximation of $y(u_\ell)$. Thus, due to the linearity of the functional $r(\cdot)$,

$$r(e - P_{\ell,C}e) = r(e) - r(P_{\ell,C}e) = r(e),$$

and, finally,

$$r(e) = \underbrace{(f + u_\ell, e)_{0,\Omega} - \langle \tilde{\sigma}_\ell, e\rangle}_{\overset{(6.3.2a)}{=} a(y(u_\ell),e)} - a(y_\ell, e) = a(e, e).$$

Relation (6.3.5) serves as a starting point for the derivation of the upper bound (6.3.4b). For the sake of convenience, let us set $\bar{e} := e - P_{\ell,C}e$. Performing element-wise integration by parts and taking advantage of the representation (4.5.6a) of the extension $\tilde{\sigma}_\ell$ of the discrete multiplier σ_ℓ, by straightforward elimination we obtain

$$\begin{aligned} r(\bar{e}) &= \sum_{T\in\mathcal{T}_\ell} (f + u_\ell, \bar{e})_{0,T} - \sum_{E\in\mathcal{E}_\ell} (\nu_E \cdot [\nabla y_\ell], \bar{e})_{0,E} - \langle \tilde{\sigma}_\ell, \bar{e}\rangle \\ &= \sum_{T\subseteq\mathcal{Z}_\ell} (f + u_\ell, \bar{e})_{0,T} - \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} (\nu_E \cdot [\nabla y_\ell], \bar{e})_{0,E} - F_{\ell,\sigma}(P_{\ell,SZ}\bar{e}), \end{aligned}$$

and, consequently,

$$|r(\bar{e})| \le |\sum_{T\subseteq\mathcal{Z}_\ell}(f+u_\ell,\bar{e})_{0,T}| + |\sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}}(\nu_E\cdot[\nabla y_\ell],\bar{e})_{0,E}| + |F_{\ell,\sigma}(P_{\ell,SZ}\bar{e})|.$$

The estimation of the first two terms on the right-hand side is based on the properties of the Clément's quasi-interpolation operator (4.1.3) that yield

$$\|\bar{e}\|_{0,T} \lesssim h_T|e|_{1,\omega_\ell^T}, \quad \|\bar{e}\|_{0,E} \lesssim h_T^{1/2}|e|_{1,\omega_\ell^E}$$

(see Figure 4.1 for the visualization of the patches ω_ℓ^T and ω_ℓ^E). We obtain

$$|\sum_{T\subseteq\mathcal{Z}_\ell}(f+u_\ell,\bar{e})_{0,T}| \le \sum_{T\subseteq\mathcal{Z}_\ell}\|f+u_\ell\|_{0,T}\|\bar{e}\|_{0,T} \lesssim \sum_{T\subseteq\mathcal{Z}_\ell}\eta_T(y)|e|_{1,\omega_\ell^T}$$

and

$$|\sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}}(\nu_E\cdot[\nabla y_\ell],\bar{e})_{0,E}| \le \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}}\|(\nu_E\cdot[\nabla y_\ell]\|_{0,E}\|\bar{e}\|_{0,E} \lesssim \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}}\eta_E(y)|e|_{1,\omega_\ell^E}.$$

To estimate the third term, we use the properties of the Scott-Zhang interpolation operator, preliminary bounding the term from above by the sum

$$\begin{aligned}|F_{\ell,\sigma}(P_{\ell,SZ}\bar{e})| \le \sum_{T\subseteq\mathcal{F}_\ell(\sigma_\ell)}\Big(\|f+u_\ell\|_{0,T}\sum_{a\in\mathcal{N}_\ell(T)\cap C_\ell}\|P_{\ell,SZ}\bar{e}(a)\varphi_\ell^{(a)}\|_{0,T}\Big)\\ + \sum_{E\in\mathcal{E}_{\mathcal{F}_\ell(\sigma_\ell)}}\|\nu_E\cdot[\nabla y_\ell]\|_{0,E}\|P_{\ell,SZ}\bar{e}(a'_E)\varphi_\ell^{(a'_E)}\|_{0,E},\\ a'_E=\mathcal{N}_\ell(E)\cap C_\ell,\ E\in\mathcal{E}_{\mathcal{F}_\ell(\sigma_\ell)}.\end{aligned}$$

The elementary properties of the nodal basis functions

$$\|\varphi_\ell^{(a)}\|_{0,T} \lesssim h_T, \quad a\in\mathcal{N}_\ell(T), \tag{6.3.6}$$

$$\|\varphi_\ell^{(a)}\|_{0,E} \lesssim h_E^{1/2},\ a\in\mathcal{N}_\ell(E) \tag{6.3.7}$$

(see, e.g., [18], Th. 3.1.2.), as well as the properties of the Scott-Zhang interpolation operator yield

$$\sum_{a\in\mathcal{N}_\ell(T)\cap C_\ell}\|P_{\ell,SZ}\bar{e}(a)\varphi_\ell^{(a)}\|_{0,T} = \sum_{a\in\mathcal{N}_\ell(T)\cap C_\ell}|P_{\ell,SZ}\bar{e}(a)|\,\|\varphi_\ell^{(a)}\|_{0,T} \overset{(6.3.6)}{\lesssim} h_T\sum_{a\in\mathcal{N}_\ell(T)\cap C_\ell}|P_{\ell,SZ}\bar{e}(a)|$$

$$\overset{\text{Prop. 4.1.7 (ii)}}{\lesssim} h_T\sum_{a\in\mathcal{N}_\ell(T)\cap C_\ell}h_{T^{(a)}}^{-1}\|\bar{e}\|_{0,T^{(a)}} \lesssim h_T\sum_{a\in\mathcal{N}_\ell(T)\cap C_\ell}|e|_{1,\omega_\ell^{T^{(a)}}}$$

and

$$\|P_{\ell,SZ}\bar{e}(a'_E)\,\varphi_\ell^{(a'_E)}\|_{0,E} = |P_{\ell,SZ}\bar{e}(a'_E)|\,\|\varphi_\ell^{(a'_E)}\|_{0,E} \overset{(6.3.7)}{\lesssim} h_E^{1/2}h_{T^{(a'_E)}}^{-1}\|\bar{e}\|_{0,T^{(a'_E)}} \lesssim h_E^{1/2}|e|_{1,\omega_\ell^{T^{(a'_E)}}}.$$

Here, $T^{(a)}$ stands for the fixed element in ω_ℓ^a which is used in the computation of the nodal coefficient $P_{\ell,SZ}\bar{e}(a)$. Hence, it follows that

$$|F_{\ell,\sigma}(P_{\ell,SZ}\bar{e})| \lesssim \sum_{T\subseteq\mathcal{F}_\ell(\sigma_\ell)} \eta_T(y)|e|_{1,\widetilde{\omega}^T} + \sum_{E\in\mathcal{E}_{\mathcal{F}_\ell(\sigma_\ell)}} \eta_E(y)|e|_{1,\omega_\ell^{T^{(a'_E)}}},$$

where

$$\widetilde{\omega}^T := \bigcup_{a\in\mathcal{N}_\ell(T)\cap\mathcal{C}_\ell} \omega_\ell^{T^{(a)}}.$$

Summarizing all the above estimates, we deduce

$$|r(\bar{e})| \lesssim \sum_{T\subseteq\mathcal{Z}_\ell} \eta_T(y)|e|_{1,\widehat{\omega}^T} + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} \eta_E(y)\ |e|_{1,\widehat{\omega}^E},$$

where

$$\widehat{\omega}^T := \begin{cases} \widetilde{\omega}^T \cup \omega_\ell^T, & \text{if } T\subseteq\mathcal{F}_\ell(\sigma_\ell) \\ \omega_\ell^T, & \text{else} \end{cases}, \quad \widehat{\omega}^E := \begin{cases} \omega_\ell^{T^{(a'_E)}} \cup \omega_\ell^E, & \text{if } E\in\mathcal{E}_{\mathcal{F}_\ell(\sigma_\ell)} \\ \omega_\ell^E, & \text{else} \end{cases}.$$

Applying the Cauchy-Schwarz inequality for sums and taking into account that the patches ω_ℓ^T and ω_ℓ^E have a finite overlap, we get

$$\begin{aligned} |r(\bar{e})| &\lesssim \left(\sum_{T\subseteq\mathcal{Z}_\ell)} \eta_T^2(y)\right)^{1/2} \left(\sum_{T\subseteq\mathcal{Z}_\ell} |e|_{1,\widehat{\omega}^T}^2\right)^{1/2} + \left(\sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} \eta_E^2(y)\right)^{1/2} \left(\sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} |e|_{1,\widehat{\omega}^E}^2\right)^{1/2} \\ &\lesssim \left(\sum_{T\subseteq\mathcal{Z}_\ell} \eta_T^2(y) + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}\cup\mathcal{E}_{\mathcal{F}_\ell(\sigma_\ell)}} \eta_E^2(y)\right)^{1/2} \left(\sum_{T\subseteq\mathcal{Z}_\ell} |e|_{1,\widehat{\omega}^T}^2 + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} |e|_{1,\widehat{\omega}^E}^2\right)^{1/2} \lesssim \eta_\ell(y)|e|_{1,\Omega}. \end{aligned}$$

Using the preceding inequality and (2.1.1b) in (6.3.5) gives (6.3.4a).
For the proof of (6.3.4b) we set $e := p(y_\ell) - p_\ell$ and obtain

$$\|e\|_{1,\Omega}^2 \lesssim a(e,e) = r(e - P_{\ell,C}e), \tag{6.3.8}$$

where the residual $r(\cdot)$ is given by

$$r(v) := (y^d - y_\ell, v)_{0,\Omega} - a(p_\ell, v) - \langle\tilde{\mu}_\ell, v\rangle,\ v\in V\ .$$

Setting $\bar{e} := e - P_{\ell,C}e$, element-wise integration by parts and the representation (4.5.6b) of the extension $\tilde{\mu}$ lead to

$$\begin{aligned} r(\bar{e}) &= \sum_{T\in\mathcal{I}_\ell} (y^d - y_\ell, \bar{e})_{0,T} - \sum_{E\in\mathcal{E}_\ell} (\nu_E\cdot[\nabla p_\ell], \bar{e})_{0,E} \ - \ \langle\tilde{\mu}_\ell, \bar{e}\rangle \\ &= \sum_{T\subseteq\mathcal{I}_\ell} (y^d - y_\ell, \bar{e})_{0,T} - \sum_{E\in\mathcal{E}_{\mathcal{I}_\ell}} (\nu_E\cdot[\nabla p_\ell], \bar{e})_{0,E} - F_{\ell,\mu}(P_{\ell,SZ}\bar{e}). \end{aligned} \tag{6.3.9}$$

The terms on the right-hand side can be estimated from above in much the same way as before resulting in

$$|r(\bar{e})| \lesssim \eta_\ell(p)|e|_{1,\Omega},$$

which together with (6.3.8) proves the assertion. □

Lemma 6.3.4. *Let the assumptions of Theorem 6.3.1 be satisfied and let $y(u_\ell)$, $p(y_\ell)$ be the auxiliary state and the auxiliary adjoint state as given by* (6.3.2)*. Further, let $e^{(1)}_{\ell,\sigma}$ and $e^{(2)}_{\ell,\mu}$ be the consistency error terms given by* (6.2.8)*. Then, there holds*

$$\|y - y(u_\ell)\|^2_{1,\Omega} \lesssim \|e_{\ell,u}\|^2_{0,\Omega} + \eta^2_\ell(y) + e^{(1)}_{\ell,\sigma}, \tag{6.3.10a}$$

$$\|p - p(y_\ell)\|^2_{1,\Omega} \lesssim \|e_{\ell,y}\|^2_{0,\Omega} + \eta^2_\ell(p) + e^{(2)}_{\ell,\mu}. \tag{6.3.10b}$$

Proof. Subtracting (6.3.2a) from (6.1.3a) yields

$$a(y - y(u_\ell), v) = (e_{\ell,u}, v)_{0,\Omega} + e_{\ell,\sigma}(v), \ \forall v \in V. \tag{6.3.11}$$

Choosing $v = y - y(u_\ell)$ and observing (2.1.1b), we obtain

$$\begin{aligned}\gamma_a\|y - y(u_\ell)\|^2_{1,\Omega} &\leq a(y - y(u_\ell), y - y(u_\ell)) \\ &= (e_{\ell,u}, y - y(u_\ell))_{0,\Omega} + e_{\ell,\sigma}(y_\ell - y(u_\ell)) + e^{(1)}_{\ell,\sigma}.\end{aligned} \tag{6.3.12}$$

The Cauchy-Schwarz inequality and the Young's inequality ($|ab| \leq \frac{\delta}{2}a^2 + \frac{1}{2\delta}b^2$, $\delta > 0$) with $\delta = \frac{\gamma_a}{2}$ gives

$$|(e_{\ell,u}, y - y(u_\ell))_{0,\Omega}| \leq \frac{\gamma_a}{4}\|y - y(u_\ell)\|^2_{0,\Omega} + \frac{1}{\gamma_a}\|e_{\ell,u}\|^2_{0,\Omega}. \tag{6.3.13}$$

Moreover, if we choose $v = y_\ell - y(u_\ell)$ in (6.3.11), we obtain

$$e_{\ell,\sigma}(y_\ell - y(u_\ell)) = (e_{\ell,u}, y(u_\ell) - y_\ell)_{0,\Omega} + a(y - y(u_\ell), y_\ell - y(u_\ell)).$$

Another application of the Cauchy-Schwarz inequality and Young's inequality yield

$$|e_{\ell,\sigma}(y_\ell - y(u_\ell))| \leq \frac{\gamma_a}{4}\|y - y(u_\ell)\|^2_{1,\Omega} + \frac{2}{\gamma_a}\|y_\ell - y(u_\ell)\|^2_{1,\Omega} + \frac{\gamma_a}{4}\ \|e_{\ell,u}\|^2_{0,\Omega}. \tag{6.3.14}$$

Using (6.3.13) and (6.3.14) in (6.3.12), we get

$$\gamma_a\|y - y(u_\ell)\|^2_{1,\Omega} \leq \frac{\gamma_a}{2}\|y - y(u_\ell)\|^2_{1,\Omega} + \left[\frac{1}{\gamma_a} + \frac{\gamma_a}{4}\right]\|e_{\ell,u}\|^2_{0,\Omega} + \frac{2}{\gamma_a}\|y_\ell - y(u_\ell)\|^2_{1,\Omega} + e^{(1)}_{\ell,\sigma},$$

Setting

$$d_1 := \frac{\gamma_a^2 + 4}{2\gamma_a^2}, \quad d_2 := \frac{4}{\gamma_a^2}, \quad d_3 := \frac{2}{\gamma_a}, \tag{6.3.15}$$

it follows that

$$\|y - y(u_\ell)\|^2_{1,\Omega} \leq d_1\,\|e_{\ell,u}\|^2_{0,\Omega} + d_2\,\|y_\ell - y(u_\ell)\|^2_{1,\Omega} + d_3\,e^{(1)}_{\ell,\sigma}, \tag{6.3.16}$$

The second term on the right-hand side in (6.3.16) can be estimated from above by (6.3.4a) which results in (6.3.10a).

The estimate (6.3.10b) can be established by using similar arguments. Subtracting (6.3.2b) from (6.1.3c) yields

$$a(p - p(y_\ell), v) = -(e_{\ell,y}, v) + e_{\ell,\mu}(v), \ \forall v \in V. \tag{6.3.17}$$

Choosing $v = p - p(y_\ell)$ and $v = p_\ell - p(y_\ell)$, we obtain

$$\gamma_a \|p - p(y_\ell)\|_{1,\Omega}^2 \leq a(p - p(y_\ell), p - p(y_\ell)) = (e_{\ell,y}, p(y_\ell) - p)_{0,\Omega} + e_{\ell,\mu}(p_\ell - p(y_\ell)) + e_{\ell,\mu}^{(2)},$$
$$e_{\ell,\mu}(p_\ell - p(y_\ell)) = (e_{\ell,y}, p_\ell - p(y_\ell))_{0,\Omega} + a(p - p(y_\ell), p_\ell - p(y_\ell)).$$

An application of the Cauchy-Schwarz inequality and Young's inequality gives

$$\|p - p(y_\ell)\|_{1,\Omega}^2 \leq d_1 \|e_{\ell,y}\|_{0,\Omega}^2 + d_2 \|p_\ell - p(y_\ell)\|_{1,\Omega}^2 + d_3 \, e_{\ell,\mu}^{(2)}, \tag{6.3.18}$$

from which (6.3.10b) can be deduced in view of (6.3.4b). □

Lemma 6.3.5. *Let the assumptions of Theorem 6.3.1 be satisfied and let η_ℓ, $osc_{\ell,rel}$, and $e_{\ell,rel}^c$ be the residual-type error estimator* (6.2.1)*, the data oscillation $osc_{\ell,rel}$* (6.2.5)*, and the consistency error term $e_{\ell,rel}^c$* (6.2.9)*. Then, it holds*

$$\|e_u\|_{0,\Omega}^2 \lesssim \eta_\ell^2 + osc_{\ell,rel}^2 + e_{\ell,rel}^c. \tag{6.3.19}$$

Proof. Combining (6.1.2d) and (6.1.3d) we obtain

$$\begin{aligned} \|e_{\ell,u}\|_{0,\Omega}^2 &= (e_{\ell,u}, u - u_\ell)_{0,\Omega} \\ &= (e_{\ell,u}, u^d - u_\ell^d)_{0,\Omega} + (e_{\ell,u}, (u - u^d) - (u_\ell - u_\ell^d))_{0,\Omega} \\ &= (e_{\ell,u}, u^d - u_\ell^d)_{0,\Omega} + \alpha^{-1}(e_{\ell,u}, p - p_\ell)_{0,\Omega}. \end{aligned} \tag{6.3.20}$$

The first term on the right-hand side in (6.3.20) can be estimated from above using the Young's inequality with $\delta = 2$

$$|(e_{\ell,u}, u^d - u_\ell^d)_{0,\Omega}| \leq \frac{1}{4}\|e_{\ell,u}\|_{0,\Omega}^2 + osc_\ell^2(u^d). \tag{6.3.21}$$

The second term can be split according to

$$(e_{\ell,u}, p - p_\ell)_{0,\Omega} = (e_{\ell,u}, p - p(y_\ell))_{0,\Omega} + (e_{\ell,u}, p(y_\ell) - p_\ell)_{0,\Omega}. \tag{6.3.22}$$

For the estimation of the first term on the right-hand side in (6.3.22) we choose $v = p - p(y_\ell)$ in (6.3.11) which gives

$$a(y - y(u_\ell), p - p(y_\ell)) = (e_{\ell,u}, p - p(y_\ell))_{0,\Omega} + e_{\ell,\sigma}(p - p(y_\ell)). \tag{6.3.23}$$

On the other hand, choosing $v = y - y(u_\ell)$ in (6.3.17) yields

$$a(p - p(y_\ell), y - y(u_\ell)) = -(e_{\ell,y}, y - y(u_\ell))_{0,\Omega} + e_{\ell,\mu}(y - y(u_\ell)). \tag{6.3.24}$$

Combining (6.3.23) and (6.3.24) and using the symmetry of $a(\cdot,\cdot)$, it follows that

$$\begin{aligned}(e_{\ell,u}, p - p(y_\ell))_{0,\Omega} &= a(y - y(u_\ell), p - p(y_\ell)) + e_{\ell,\sigma}(p(y_\ell) - p)\\ &= -(e_{\ell,y}, y - y(u_\ell))_{0,\Omega} + e_{\ell,\mu}(y - y(u_\ell)) + e_{\ell,\sigma}(p(y_\ell) - p)\\ &= -(e_{\ell,y}, y - y(u_\ell))_{0,\Omega} + e^{(1)}_{\ell,\mu} + e^{(2)}_{\ell,\sigma} + e_{\ell,\mu}(y_\ell - y(u_\ell)) + e_{\ell,\sigma}(p(y_\ell) - p_\ell).\end{aligned}$$

Now, choosing $v = y_\ell - y(u_\ell)$ in (6.3.17) and $v = p(y_\ell) - p_\ell$ in (6.3.11), we find

$$\begin{aligned}e_{\ell,\mu}(y_\ell - y(u_\ell)) &= (e_{\ell,y}, y_\ell - y(u_\ell))_{0,\Omega} + a(p - p(y_\ell), y_\ell - y(u_\ell)),\\ e_{\ell,\sigma}(p(y_\ell) - p_\ell) &= -(e_{\ell,u}, p(y_\ell) - p_\ell)_{0,\Omega} + a(y - y(u_\ell), p(y_\ell) - p_\ell),\end{aligned}$$

and hence,

$$\begin{aligned}(e_{\ell,u}, p - p(y_\ell))_{0,\Omega} &= -\|e_{\ell,y}\|^2_{0,\Omega} + e^{(1)}_{\ell,\mu} + e^{(2)}_{\ell,\sigma} +\\ &+ a(p - p(y_\ell), y_\ell - y(u_\ell)) + a(y - y(u_\ell), p(y_\ell) - p_\ell) - (e_{\ell,u}, p(y_\ell) - p_\ell)_{0,\Omega}.\end{aligned} \tag{6.3.25}$$

Using (6.3.25) in (6.3.22) results in

$$\begin{aligned}(e_{\ell,u}, p - p_\ell)_{0,\Omega} &= -\|e_{\ell,y}\|^2_{0,\Omega} + e^{(1)}_{\ell,\mu} + e^{(2)}_{\ell,\sigma}\\ &+ \underbrace{a(p - p(y_\ell), y_\ell - y(u_\ell))}_{=:\,[I]} + \underbrace{a(y - y(u_\ell), p(y_\ell) - p_\ell)}_{=:\,[II]}.\end{aligned} \tag{6.3.26}$$

Young's inequality gives

$$\begin{aligned}[I] &\le |p - p(y_\ell)|_{1,\Omega}|y_\ell - y(u_\ell)|_{1,\Omega} \le \|p - p(y_\ell)\|_{1,\Omega}\|y_\ell - y(u_\ell)\|_{1,\Omega}\\ &\le \frac{1}{2\delta}\|p - p(y_\ell)\|^2_{1,\Omega} + \frac{\delta}{2}\|y_\ell - y(u_\ell)\|^2_{1,\Omega}, \quad \forall \delta > 0.\end{aligned}$$

Using (6.3.18) and choosing $\delta = d_1/2$ (cf. (6.3.15)), we get

$$[I] \le \|e_{\ell,y}\|^2_{0,\Omega} + \frac{d_2}{d_1}\|p_\ell - p(y_\ell)\|^2_{1,\Omega} + \frac{d_3}{d_1}\, e^{(2)}_{\ell,\mu} + \frac{d_1}{4}\|y_\ell - y(u_\ell)\|^2_{1,\Omega}. \tag{6.3.27}$$

Similarly,

$$[II] \le \frac{1}{2\delta}\|y - y(u_\ell)\|^2_{1,\Omega} + \frac{\delta}{2}\|p_\ell - p(y_\ell)\|^2_{1,\Omega}, \quad \forall \delta > 0.$$

Observing (6.3.16), we choose $\delta = 2d_1/\alpha$ and obtain

$$[II] \le \frac{\alpha}{4}\|e_{\ell,u}\|^2_{0,\Omega} + \frac{\alpha d_2}{4d_1}\|y_\ell - y(u_\ell)\|^2_{1,\Omega} + \frac{\alpha d_3}{4d_1}e^{(1)}_{\ell,\sigma} + \frac{d_1}{\alpha}\|p_\ell - p(y_\ell)\|^2_{1,\Omega}. \tag{6.3.28}$$

Using (6.3.21) and (6.3.26) - (6.3.28) in (6.3.20), it follows that

$$\|e_{\ell,u}\|^2_{0,\Omega} \lesssim \|p_\ell - p(y_\ell)\|^2_{1,\Omega} + \|y_\ell - y(u_\ell)\|^2_{1,\Omega} + osc^2_{\ell,rel} + e^c_{\ell,rel}. \tag{6.3.29}$$

The assertion (6.3.19) follows from (6.3.29) by taking (6.3.4a), (6.3.4b) from Lemma 6.3.3 into account. □

Proof of Theorem 6.3.1. The estimate (6.3.1) follows from the preceding Lemmas 6.3.3, 6.3.4, and 6.3.5 in view of

$$e_{\ell,y} = y - y(u_\ell) + y(u_\ell) - y_\ell,$$
$$e_{\ell,p} = p - p(y_\ell) + p(y_\ell) - p_\ell.$$

and the Young's inequality, which imply

$$\begin{aligned}
\|e_{\ell,y}\|_{1,\Omega}^2 &\leq \frac{3}{2}\|y - y(u_\ell)\|_{1,\Omega}^2 + \frac{3}{2}\|y(u_\ell) - y_\ell\|_{1,\Omega}^2 \qquad (6.3.30)\\
&\overset{\substack{Lem.6.3.3\\Lem.6.3.4}}{\lesssim} \|e_{\ell,u}\|_{0,\Omega}^2 + \eta_\ell^2(y) + e_{\ell,\sigma}^{(1)} \overset{Lem.6.3.5}{\lesssim} \eta_\ell^2 + osc_{\ell,rel}^2 + e_{\ell,rel}^c\,,\\
\|e_{\ell,p}\|_{1,\Omega}^2 &\leq \frac{3}{2}\|p - p(y_\ell)\|_{1,\Omega}^2 + \frac{3}{2}\|p(y_\ell) - p_\ell\|_{1,\Omega}^2\\
&\overset{\substack{Lem.6.3.3\\Lem.6.3.4}}{\lesssim} \|e_{\ell,y}\|_{0,\Omega}^2 + \eta_\ell^2(p) + e_{\ell,\mu}^{(2)} \overset{(6.3.30)}{\lesssim} \eta_\ell^2 + osc_{\ell,rel}^2 + e_{\ell,rel}^c\,.
\end{aligned}$$

□

6.4 Proof of the efficiency estimate

Theorem 6.4.1. *Let (y, σ, u, p, μ) and $(y_\ell, \sigma_\ell, u_\ell, p_\ell, \mu_\ell)$ be solutions of* (6.1.3) *and* (6.1.2)*, respectively, and let η_ℓ, $osc_{\ell,eff}$, and $e_{\ell,eff}^c$ be the error estimator, the data oscillation, and the consistency error as given by* (6.2.1), (6.2.6) *and* (6.2.10)*. Then, it holds*

$$\eta_\ell^2 - osc_{\ell,eff}^2 - e_{\ell,eff}^c \lesssim |||e_\ell|||^2\,.$$

The proof of Theorem 6.4.1 will be provided by the subsequent two lemmas, taking into account the following well-known properties (see, e.g., [61]) of the element bubble functions

$$\|q_\ell\|_{0,T}^2 \lesssim (q_\ell, q_\ell b_T)_{0,T}, \quad q_\ell \in P_1(T)\,, \qquad (6.4.1a)$$
$$\|q_\ell b_T\|_{0,T} \lesssim \|q_\ell\|_{0,T}, \quad q_\ell \in P_1(T)\,, \qquad (6.4.1b)$$
$$h_T^{-1}\|q_\ell\|_{0,T} \lesssim |q_\ell b_T|_{1,T} \lesssim h_T^{-1}\|q_\ell\|_{0,T}, \quad q_\ell \in P_1(T)\,, \qquad (6.4.1c)$$

and of the edge bubble functions

$$\|q_\ell\|_{0,E}^2 \lesssim (q_\ell, q_\ell b_E)_{0,E}, \quad q_\ell \in P_1(E)\,, \qquad (6.4.2a)$$
$$\|q_\ell b_E\|_{0,\omega_E} \lesssim h_E^{1/2}\|q_\ell\|_{0,E}, \quad q_\ell \in P_0(E)\,, \qquad (6.4.2b)$$
$$h_E^{-1/2}\|q_\ell\|_{0,E} \lesssim |q_\ell b_E|_{1,\omega_E} \lesssim h_E^{-1/2}\|q_\ell\|_{0,E}, \quad q_\ell \in P_0(E)\,. \qquad (6.4.2c)$$

Lemma 6.4.2. *Let the assumptions of Theorem 6.4.1 be satisfied and let $\eta_T(y)$, $\eta_T(p)$, $osc_T(f)$, $osc_T(y^d)$, e_T^σ, and e_T^μ be the element residuals* (6.2.3)*, the data oscillations* (6.2.7)*, and the consistency error terms defined in* (6.2.11)*, respectively. Then, for all $T \subseteq \mathcal{Z}_\ell$ it holds*

$$\eta_T(y) \lesssim \|e_{\ell,y}\|_{1,T} + h_T\|e_{\ell,u}\|_{0,T} + osc_T(f) + e_T^\sigma, \tag{6.4.3a}$$

whereas for all $T \subseteq \mathcal{I}_\ell$ we have

$$\eta_T(p) \lesssim \|e_{\ell,p}\|_{1,T} + h_T\|e_{\ell,y}\|_{0,T} + \mathrm{osc}_T(y^d) + e_T^\mu. \tag{6.4.3b}$$

Proof. Setting $\psi_T^\sigma = (f_\ell + u_\ell)b_T \in C(\overline{\Omega}) \cap V_{T,0}$, in view of (6.4.1a), $\Delta y_\ell|_T = 0$, and Green's formula, we obtain

$$\begin{aligned} h_T^2\ \|f_\ell + u_\ell\|_{0,T}^2 &\lesssim h_T^2\ (f_\ell + u_\ell, \psi_T^\sigma)_{0,T}\ =\ h_T^2\ (f_\ell + u_\ell + \Delta y_\ell, \psi_T^\sigma)_{0,T} \\ &= h_T^2 \underbrace{\int_{\partial T} \nu_{\partial T} \cdot \nabla y_\ell\ \psi_T^\sigma\ ds}_{=0,\ as\ \psi_T^\sigma = 0\ on\ \partial T} - h_T^2\ a_T(y_\ell, \psi_T^\sigma) + h_T^2\ (f_\ell + u_\ell, \psi_T^\sigma)_{0,T}\,. \end{aligned} \tag{6.4.4}$$

On the other hand, since ψ_T^σ is an admissible test function in (6.1.3a), we have

$$a_T(y, \psi_T^\sigma) - (f + u, \psi_T^\sigma)_{0,T} + \langle \sigma, \psi_T^\sigma \rangle = 0. \tag{6.4.5}$$

Using (6.4.5) in (6.4.4), it follows that

$$\begin{aligned} h_T^2\ \|f_\ell + u_\ell\|_{0,T}^2 &\lesssim h_T^2\,[a_T(y, \psi_T^\sigma) - (f + u, \psi_T^\sigma)_{0,T} + \langle \sigma, \psi_T^\sigma \rangle] \\ &- h_T^2\,[a_T(y_\ell, \psi_T^\sigma) - (f_\ell + u_\ell, \psi_T^\sigma)_{0,T}] \\ &= h_T^2\,[a_T(y - y_\ell, \psi_T^\sigma) - (f - f_\ell, \psi_T^\sigma)_{0,T} - (u - u_\ell, \psi_T^\sigma)_{0,T} + \langle \sigma, \psi_T^\sigma \rangle] \\ &\le h_T^2\,[|e_{\ell,y}|_{1,T}|\psi_T^\sigma|_{1,T} + \|e_{\ell,u}\|_{0,T}\|\psi_T^\sigma\|_{0,T} + \|f - f_\ell\|_{0,T}\|\psi_T^\sigma\|_{0,T} + e_T^\sigma|\psi_T^\sigma|_{1,T}]\,. \end{aligned} \tag{6.4.6}$$

In view of (6.4.1b) and (6.4.1c), it holds

$$\begin{aligned} &h_T^{-1}\|f_\ell + u_\ell\|_{0,T} \lesssim |\psi_T^\sigma|_{1,T} = |b_T \underbrace{(f_\ell + u_\ell)}_{\in P_1(T)}|_{1,T} \lesssim h_T^{-1}\|f_\ell + u_\ell\|_{0,T}, \\ &\|\psi_T^\sigma\|_{0,T} \lesssim \|f_\ell + u_\ell\|_{0,T}. \end{aligned} \tag{6.4.7}$$

Now, using (6.4.7) in (6.4.6), we get

$$h_T^2\|f_\ell + u_\ell\|_{0,T}^2 \lesssim h_T\|f_\ell + u_\ell\|_{0,T}\Big[\|e_{\ell,y}\|_{1,T} + h_T\|e_{\ell,u}\|_{0,T} + osc_T(f) + e_T^\sigma\Big].$$

Combining the preceding estimate with $\eta_T(y) \le h_T\|f_\ell + u_\ell\|_{0,T} + osc_T(f)$ yields (6.4.3a).

For the proof of (6.4.3b) we choose $\psi_T^\mu := (y_\ell^d - y_\ell)b_T \in C(\overline{\Omega}) \cap V_{T,0}$. Again, taking advantage of (6.4.1a), $\Delta p_h|_T = 0$, and Green's formula, we get

$$h_T^2\|y_\ell^d - y_\ell\|_{0,T}^2 \lesssim h_T^2(y_\ell^d - y_\ell, \psi_T^\mu)_{0,T} = -h_T^2 a_T(p_\ell, \psi_T^\mu) + h_T^2(y_\ell^d - y_\ell, \psi_T^\mu)_{0,T}. \tag{6.4.8}$$

Since ψ_T^μ is admissible in (6.1.3c), we have

$$a_T(p, \psi_T^\mu) - (y^d - y, \psi_T^\mu)_{0,T} + \langle \mu, \psi_T^\mu \rangle = 0.$$

Adding this expression to the right-hand side of (6.4.8), we obtain

$$\begin{aligned} h_T^2 \|y_\ell^d - y_\ell\|_{0,T}^2 &\lesssim h_T^2 \left[a_T(p - p_\ell, \psi_T^\mu) + (y - y_\ell, \psi_T^\mu)_{0,T} - (y^d - y_\ell^d, \psi_T^\mu)_{0,T} + \langle \mu, \psi_T^\mu \rangle \right] \\ &\lesssim h_T^2 |e_p|_{1,T} |\psi_T^\mu|_{1,T} + h_T^2 \|e_y\|_{0,T} \|\psi_T^\mu\|_{0,T} + h_T \, osc_T(y^d) \|\psi_T^\mu\|_{0,T} + h_T^2 e_T^\mu |\psi_T^\mu|_{1,T}. \end{aligned}$$

Finally, an application of (6.4.1b), (6.4.1c) leads to

$$\eta_T(p) \leq h_T \|y_\ell^d - y_\ell\|_{0,T} + osc_T(y^d) \lesssim \|e_{\ell,p}\|_{1,T} + h_T \|e_{\ell,y}\|_{0,T} \, , + osc_T(y^d) + e_T^\mu$$

which implies (6.4.3b). □

Lemma 6.4.3. *Let the assumptions of Theorem 6.4.1 be satisfied and let $\eta_T(y)$, $\eta_T(p)$, $e_{\omega_E}^\sigma$ and $e_{\omega_E}^\mu$ be the edge residuals* (6.2.4) *and the consistency error terms defined in* (6.2.11)*, respectively. Further, for $E = T_+ \cap T_-$, $T_\pm \in \mathcal{T}_\ell$ let*

$$\eta_{\omega_E}(y) := \eta_{T_+}(y) + \eta_{T_-}(y), \quad osc_{\omega_E}(f) := osc_{T_+}(f) + osc_{T_-}(f), \tag{6.4.9a}$$

$$\eta_{\omega_E}(p) := \eta_{T_+}(p) + \eta_{T_-}(p), \quad osc_{\omega_E}(y^d) := osc_{T_+}(y^d) + osc_{T_-}(y^d). \tag{6.4.9b}$$

Then, for all $E \in \mathcal{E}_{\mathcal{Z}_\ell}$ we have

$$\eta_E(y) \lesssim \|e_{\ell,y}\|_{1,\omega_E} + h_E \|e_{\ell,u}\|_{0,\omega_E} + osc_{\omega_E}(f) + \eta_{\omega_E}(y) + e_{\omega_E}^\sigma, \tag{6.4.10a}$$

whereas for all $E \in \mathcal{E}_{\mathcal{I}_\ell}$ it holds

$$\eta_E(p) \lesssim \|e_{\ell,p}\|_{1,\omega_E} + h_E \|e_{\ell,y}\|_{0,\omega_E} + osc_{\omega_E}(y^d) + \eta_{\omega_E}(p) + e_{\omega_E}^\mu \, . \tag{6.4.10b}$$

Proof. For $E \in \mathcal{E}_{\mathcal{Z}_\ell}$ we set $\psi_E^\sigma := \nu_E \cdot [\nabla y_\ell] b_E \in C(\overline{\Omega}) \cup V_{\omega_E,0}$. Note that in our setting $\nu_E \cdot [\nabla y_\ell] \in P_0(E)$. Then, (6.4.2a) implies

$$\begin{aligned} \eta_E^2(y) = h_E \|\nu_E \cdot [\nabla y_\ell]\|_{0,E}^2 &\lesssim h_E (\nu_E \cdot [\nabla y_\ell], \psi_E^\sigma)_{0,E} \\ &= h_E (\nu_{\partial T_+} \cdot \nabla y_\ell|_{T_+}, \psi_E^\sigma)_{0,E} + h_E (\nu_{\partial T_-} \cdot \nabla y_\ell|_{T_-}, \psi_E^\sigma)_{0,E}. \end{aligned} \tag{6.4.11}$$

Using Green's formula, $\Delta y_\ell|_{T_\pm} = 0$, and the fact that $\psi_E^\sigma|_{T'} = 0$, $E' \in \partial T_\pm \setminus \{E\}$ (property inherited from b_E) yields

$$a_{T_\pm}(y_\ell, \psi_E^\sigma) = (\nabla y_\ell, \nabla \psi_E^\sigma)_{0,T_\pm} = (\nu_{\partial T_\pm} \cdot \nabla y_\ell|_{T_\pm}, \psi_E^\sigma)_{0,E}. \tag{6.4.12}$$

Using (6.4.12) in (6.4.11) gives

$$\eta_E^2(y) \lesssim h_E \; a_{\omega_E}(y_\ell, \psi_E^\sigma). \tag{6.4.13}$$

Since ψ_E^σ is an admissible test function in (6.1.3a), we get

$$a_{\omega_E}(y, \psi_E^\sigma) \; - \; (f + u, \psi_E^\sigma)_{0,\omega_E} \; + \; \langle \sigma, \psi_E^\sigma \rangle = 0 \; . \tag{6.4.14}$$

Combining (6.4.14) and (6.4.13), we obtain

$$\begin{aligned}\eta_E^2(y) &\lesssim h_E a(y_\ell - y, \psi_E^\sigma) + h_E(f_\ell + u_\ell, \psi_E^\sigma)_{0,\omega_E} \\ &\quad + h_E(f - f_\ell, \psi_E^\sigma)_{0,\omega_E} + h_E(u - u_\ell, \psi_E^\sigma)_{0,\omega_E} - h_E\langle\sigma, \psi_E^\sigma\rangle \\ &\le h_E|y - y_\ell|_{1,\omega_E}|\psi_E^\sigma|_{1,\omega_E} + h_E\,\|f_\ell + u_\ell\|_{0,\omega_E}\|\psi_E^\sigma\|_{0,\omega_E} \\ &\quad + h_E\|u - u_\ell\|_{0,\omega_E}\|\psi_E^\sigma\|_{0,\omega_E} + h_E\,\|f - f_\ell\|_{0,\omega_E}\|\psi_E^\sigma\|_{0,\omega_E} + h_E\, e_{\omega_E}^\sigma|\psi_E^\sigma|_{1,\omega_E}.\end{aligned} \tag{6.4.15}$$

Moreover, (6.4.2b) and (6.4.2c) imply

$$\begin{aligned}&h_E^{-1/2}\|\nu_E \cdot [\nabla y_\ell]\|_{0,E} \lesssim |\psi_E^\sigma|_{1,\omega_E} = |b_E\nu_E \cdot [\nabla y_\ell]|_{1,\omega_E} \lesssim h_E^{-1/2}\|\nu_E \cdot [\nabla y_\ell]\|_{0,E}, \\ &\|\psi_E^\sigma\|_{0,\omega_E} \lesssim h_E^{1/2}\|\nu_E \cdot [\nabla y_\ell]\|_{0,E}.\end{aligned} \tag{6.4.16}$$

Using (6.4.16) in (6.4.15) yields

$$\eta_E(y) \lesssim \|e_{\ell,y}\|_{1,\omega_E} + h_E\|e_{\ell,u}\|_{0,\omega_E} + h_E\|f_\ell + u_\ell\|_{0,\omega_E} + osc_{\omega_E}(f) + e_{\omega_E}^\sigma.$$

Due to the shape regularity of the triangulation, for $E \in \mathcal{E}_\ell(T)$ we have $h_E \lesssim h_T \lesssim h_E$ and hence,

$$\begin{aligned}&h_E\|f_\ell + u_\ell\|_{0,\omega_E} \le h_E\|f_\ell + u_\ell\|_{0,T_+} + h_E\|f_\ell + u_\ell\|_{0,T_-} \\ &\lesssim h_{T_+}\|f_\ell + u_\ell\|_{0,T_+} + h_{T_-}\|f_\ell + u_\ell\|_{0,T_-} = \eta_{\omega_E}(y).\end{aligned}$$

The preceding two estimates result in (6.4.10a) . The assertion (6.4.10b) can be verified by similar arguments. □

Proof of Theorem 6.4.1. Due to (6.2.1) and (6.2.2), we have

$$\eta_\ell^2 = \sum_{T\subseteq\mathcal{Z}_\ell}\eta_T^2(y) + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}}\eta_E^2(y) + \sum_{T\subseteq\mathcal{I}_\ell}\eta_T^2(p) + \sum_{E\in\mathcal{E}_{\mathcal{I}_\ell}}\eta_E^2(p).$$

The element and edge residuals on the right-hand side can be estimated by means of (6.4.3), (6.4.10), and the Young's inequality:

$$\begin{aligned}\eta_\ell^2 &\lesssim \|e_{\ell,y}\|_{1,\Omega}^2 + \|e_{\ell,p}\|_{1,\Omega}^2 + \|e_{\ell,u}\|_{0,\Omega}^2 + osc_{\mathcal{Z}_\ell}^2(f) + osc_{\mathcal{I}_\ell}^2(y^d) + e_{\ell,eff}^c \\ &\quad + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} osc_{\omega_E}^2(f) + \sum_{E\in\mathcal{E}_{\mathcal{I}_\ell}} osc_{\omega_E}^2(y^d) + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}}\eta_{\omega_E}^2(y) + \sum_{E\in\mathcal{E}_{\mathcal{I}_\ell}}\eta_{\omega_E}^2(p).\end{aligned}$$

Due to (4.3.1) and (6.4.9), the residuals $\eta_{\omega_E}(y)$, $\eta_{\omega_E}(p)$ can be estimated by (6.4.3). Taking advantage of the finite overlap of the patches, we get

$$\eta_\ell^2 \lesssim \|e_{y\ell,}\|_{1,\Omega}^2 + \|e_{\ell,p}\|_{1,\Omega}^2 + \|e_{\ell,u}\|_{0,\Omega}^2 + osc_{\mathcal{Z}_\ell}^2(f) + osc_{\mathcal{I}_\ell}^2(y^d) + e_{\ell,eff}^c.$$

□

6.5 Discussion of the a posteriori error estimate

In this section, we will clarify the meaning of the a posteriori error estimate (6.1.1) in view of the non-uniqueness of the continuous stationary points, discuss the consistency error terms $e^c_{\ell,rel}$, $e^c_{\ell,eff}$, and introduce a simplified form of the consistency error term $e^c_{\ell,eff}$.

6.5.1 Simplified form of the consistency error $e^c_{\ell,eff}$

It is possible to eliminate the functions ψ^σ_T, ψ^μ_T, ψ^σ_E, and ψ^μ_E from the expression of $e^c_{\ell,eff}$. We have

$$e^c_{\ell,eff} \leq \left[\sum_{T\subseteq\mathcal{Z}_\ell} |e^\sigma_T| \; + \sum_{T\subseteq\mathcal{I}_\ell} |e^\mu_T| \; + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} |e^\sigma_{\omega_E}| \; + \sum_{E\in\mathcal{E}_{\mathcal{I}_\ell}} |e^\mu_{\omega_E}| \right]^2 . \tag{6.5.1}$$

On one hand, using (2.3.5) in (6.5.1), we can obtain an upper estimate of $e^c_{\ell,eff}$ in terms of local $H^{-1}-$norms:

$$e^c_{\ell,eff} \lesssim \left[\sum_{T\subseteq\mathcal{Z}_\ell} \|\sigma\|_{-1,T} + \sum_{T\subseteq\mathcal{I}_\ell} \|\mu\|_{-1,T} + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} \|\sigma\|_{-1,\omega_E} + \sum_{E\in\mathcal{E}_{\mathcal{I}_\ell}} \|\mu\|_{-1,\omega_E} \right]^2 . \tag{6.5.2}$$

On the other hand, if σ, $\mu \in L^2(\Omega)$, applying the Hölder inequality in the nominator and the Friedrichs inequality in the denominator of each summand in (6.5.1), we get the following estimate in terms of local L^2-norms

$$e^c_{\ell,eff} \lesssim \left[\sum_{T\subseteq\mathcal{Z}_\ell} h_T\|\sigma\|_{0,T} \; + \sum_{T\subseteq\mathcal{I}_\ell} h_T\|\mu\|_{0,T} \; + \sum_{E\in\mathcal{E}_{\mathcal{Z}_\ell}} h_E\|\sigma\|_{0,\omega_E} \; + \sum_{E\in\mathcal{E}_{\mathcal{I}_\ell}} h_E\|\mu\|_{0,\omega_E} \right]^2 .$$

which, in view of finite overlap of patches, can obviously be reduced to

$$e^c_{\ell,eff} \;\lesssim\; e^{c,L^2}_{\ell,eff} := \sum_{T\subseteq\mathcal{Z}_\ell} h^2_T\|\sigma\|^2_{0,T} \; + \sum_{T\subseteq\mathcal{I}_\ell} h^2_T\|\mu\|^2_{0,T}. \tag{6.5.3}$$

The latter estimate will be used as the basis for the heuristic approximation of $e^c_{\ell,eff}$ suggested in the following section.

6.5.2 Interpretation of the error estimate in view of the non-uniqueness of solutions

In this paragraph, we will employ the notation U, U_ℓ for stationary points (cf. Section 5.3) and $D := (f, y^d, u^d)$, $D_\ell := (f_\ell, y^d_\ell, u^d_\ell)$ for the data of the continuous and discrete optimal control problems.

For the sake of brevity, the information about the dependence on U, U_ℓ and D, D_ℓ was omitted in the notation used for the components of the a posteriori error estimate (6.1.1). In order to understand the meaning of the estimate from the perspective of non-uniqueness of stationary points, we have to add an explicit record of the dependence on U, U_ℓ, D, and D_ℓ to all of its components. In particular, we have

$$\begin{aligned} &e_\ell = e_\ell(U,U_\ell), && e^c_{\ell,eff} = e^c_{\ell,eff}(U,U_\ell), && e^c_{\ell,rel} = e^c_{\ell,rel}(U,U_\ell),\\ &\eta_\ell = \eta_\ell(U_\ell,D_\ell), && osc_{\ell,eff} = osc_{\ell,eff}(D,D_\ell), && osc_{\ell,rel} = osc_{\ell,rel}(D,D_\ell), \end{aligned}$$

whence, the estimate (6.1.1) can be rewritten as

$$\begin{aligned} \eta^2_\ell(U_\ell,D_\ell) - osc^2_{\ell,eff}(D,D_\ell) - e^c_{\ell,eff}(U,U_\ell) &\lesssim |||e_\ell(U,U_\ell)|||^2 \\ &\lesssim \eta^2_\ell(U_\ell,D_\ell) + osc^2_{\ell,rel}(D,D_\ell) + e^c_{\ell,rel}(U,U_\ell). \end{aligned} \tag{6.5.4}$$

The arguments U_ℓ, D, and D_ℓ are known and fixed, whereas U can represent *any* (usually unknown) solution of the continuous system (6.1.3). The error estimate thus contains *incomputable* components dependent on U: these are the consistency errors terms $e^c_{\ell,eff}(U,U_\ell)$ and $e^c_{\ell,rel}(U,U_\ell)$. Hence, realistically, (6.5.4) is only suited for $U = U^*$ with

$$|e^c_{\ell,rel}(U^*,U_\ell)|/\eta^2_\ell(U_\ell,D_\ell) \ll 1, \quad |e^c_{\ell,eff}(U^*,U_\ell)|/\eta^2_\ell(U_\ell,D_\ell) \ll 1, \tag{6.5.5}$$

so that the unknown quantities $e^c_{\ell,rel}(U^*,U_\ell)$, $e^c_{\ell,eff}(U^*,U_\ell)$ can be genuinely neglected in the expression of the estimate. The latter can be expected only if U_ℓ is "close" to U^* as we will see in the next paragraph. Then, given the data oscillation terms are small, the computable error estimator $\eta_\ell(U_\ell,D_\ell)$ will provide us with the essential information about the error $|||e_\ell(U^*,U_\ell)|||^2$. Thus, when applying the a posteriori error estimate in practice, the following is implicitly assumed.

Assumption 6.5.1. *In the context of discrete stationary points computed on sequences of finite element triangulations $\{\mathcal{T}_\ell\}_\mathbb{N}$, we assume that the sequence of discrete stationary points $\{U_\ell\}_\mathbb{N}$ converges to a continuous stationary point U^* in such a way that*

$$\text{there exists } \bar\ell \in \mathbb{N} \text{ such that } (6.5.5) \text{ is satisfied for } \ell \geq \bar\ell. \tag{6.5.6}$$

Note that if $\{\mathcal{T}_\ell\}_\mathbb{N}$ is a sequence of uniformly refined meshes, Assumption 6.5.1 is to some extent supported by the convergence result of Theorem 5.1.4. The situation is more subtle in case of adaptively refined meshes: no convergence result for adaptive solutions is available so far. In the numerical experiments, the magnitude of the consistency errors will be compared with that of the error estimator for examples with known analytical solution. Theoretical justification of (6.5.6) for sequences of uniform and adaptive meshes is left out of scope of this work. Note, however, that if (6.5.5) are not satisfied, the applicability of the estimate (6.5.4) should be put into question. We will discuss the consistency errors in more detail in the next paragraph.

Remark 6.5.2. Considering the issue of the non-uniqueness of the solution, the following can be stated: the a posteriori error estimate (6.1.1) should be viewed and used as an estimate of the difference between U_ℓ and the continuous stationary point approximated by U_ℓ (the continuous stationary point with the ratios (6.5.5) of possibly smallest magnitude). In other words, once we are converging towards a solution, the non-uniqueness is irrelevant with respect to our a posteriori error estimate: we are basically in a situation equivalent to solving a problem with a unique solution. However, it has to be emphasized that convergence to a solution does not automatically imply the consistency errors to be small compared to the error estimator.

6.5.3 Consistency errors and local properties of solutions

We can expect consistency errors $e^c_{\ell,rel}$, $e^c_{\ell,eff}$ to be small comparing to η^2_ℓ if two conditions are satisfied simultaneously:

- (*i*) the continuous subsets of Ω are well approximated by their discrete prototypes;
- (*ii*) the local (complementarity) properties of the functions and functionals involved lead to mutual cancellation of parts of their dual products;
- (*iii*) the solution functions and functionals are well-approximated by their discrete equivalents.

Condition (*i*) can be rephrased in terms of the following assumption.

Assumption 6.5.3. *For $\ell \to \infty$, there holds*

$$\|\chi_{\mathcal{Z}_\ell} - \chi_{\mathcal{Z}}\|_{L^1(\Omega)} \;\to\; 0 \quad \textit{and} \quad \|\chi_{\mathcal{I}_\ell} - \chi_{\mathcal{I}}\|_{L^1(\Omega)} \;\to\; 0.$$

Again, though plausible for the case of uniform mesh refinement, the validity of this assumption in case of adaptively generated meshes will have to be verified in the numerical experiments.

Local properties of the discrete and continuous solution functions mentioned in (*ii*) are given by (6.1.2b), (6.1.2e), (6.1.2f), (6.1.3b), (6.1.3e), and (6.1.3f). These properties are essential for the understanding of the behavior of the consistency error terms. Let us look, for instance, at the component $e^{(2)}_{\ell,\sigma}$ of $e^c_{\ell,rel}$. As $e_{\ell,\sigma} = 0$ in $\mathcal{Z} \cap \mathring{\mathcal{Z}}_\ell$ and $e_{\ell,p} = 0$ in $\mathcal{C} \cap \mathcal{C}_\ell$, only quantities from the set

$$\Omega^{(2)}_{\ell,\sigma} := \Omega \setminus ((\mathcal{Z} \cap \mathring{\mathcal{Z}}_\ell) \cup (\mathcal{C} \cap \mathcal{C}_\ell))$$

can possibly contribute to $e^{(2)}_{\ell,\sigma}$. If additionally $\hat{\sigma}_\ell$, p_ℓ approximate σ, p in some sense (cf. condition (*iii*)), due to $e_{\ell,\sigma} \approx 0$ in $\mathcal{C} \cap (\mathcal{C}_\ell \cup \mathcal{F}_\ell(\sigma_\ell))$ and $e_{\ell,p} \approx 0$ in $\mathcal{Z} \cap \mathcal{Z}_\ell$, the essential contributions to $e^{(2)}_{\ell,\sigma}$ could only come from the set

$$\widetilde{\Omega}^{(2)}_{\ell,\sigma} := \Omega^{(2)}_{\ell,\sigma} \setminus \mathcal{F}_\ell(\sigma_\ell).$$

The other terms of $e^c_{\ell,rel}$ can be analyzed similarly. What concerns $e^c_{\ell,eff}$, in view of (6.5.2) and (4.3.1), only the part of σ associated with $\mathcal{C} \cap \mathcal{Z}_\ell$ and of μ with $\mathcal{A} \cap \mathcal{I}_\ell$ will contribute. Thus, when $\mathcal{Z} \cap \mathcal{Z}_\ell$ and $\mathcal{I} \cap \mathcal{I}_\ell$ are small, we may expect the quantities $e^c_{\ell,rel}$ and $e^c_{\ell,eff}$ to be small compared to the error estimator η_ℓ that lives on significantly larger sets $\mathcal{Z}_\ell$ and $\mathcal{I}_\ell$.

Remark 6.5.4. Unlike the case of a posteriori error analysis for the state constraint optimal control problem suggested in [43], the value of $e^c_{\ell,rel}$ depends only on the continuous and discrete state and adjoint state functions, and not on the auxiliary functions $y(u_\ell)$ and $p(y_\ell)$. We are able to eliminate the auxiliary functions from the error estimate as the terms $e_{\ell,\sigma}(y(u_\ell) - y_\ell)$, $e_{\ell,\sigma}(p_\ell - p(y_\ell))$, $e_{\ell,\mu}(y(u_\ell) - y_\ell)$, and $e_{\ell,\mu}(p(y_\ell) - p_\ell)$ can be controlled by the error estimator η_ℓ. In view of (ii), this is an advantage of the error estimate as the auxiliary functions $y(u_\ell)$, $p(y_\ell)$ do not inherit any local properties of y, p and y_ℓ, p_ℓ. Moreover, being independent on the auxiliary states, $e^c_{\ell,rel}$ can be computed for those examples where the analytical solution of the continuos problem is known.

Remark 6.5.5. Due to the lack of substantial dependence of the computable parts of the error estimate on the type of stationary points, the influence of the latter on the performance of the error estimate has not been investigated in this work. However, it is not difficult to see that $e^c_{\ell,rel}$ and $e^c_{\ell,eff}$ are the only parts of the estimator sensitive to the type of the discrete and continuous stationarity points. A more detailed analysis of the consistency errors could make use of this additional information.

6.6 Approximation of the consistency errors

In this section, we will provide fully computable heuristic approximations for the reliability and efficiency consistency errors $e^c_{\ell,rel}$, $e^c_{\ell,eff}$.

6.6.1 Approximation of characteristic functions

We will follow the approach suggested in [32, 34, 47] for the case of adaptive finite element approximations of control and/or state constrained optimally controlled second order elliptic boundary value problems. First, we will provide approximations of the characteristic functions $\chi_{\mathcal{A}}$ and $\chi_{\mathcal{Z}}$ of the active set $\mathcal{A}$ and the zero set $\mathcal{Z}$ by means of the available finite element solutions. The original method of [47] will be improved by *the local mesh size* $\bar{h}_\ell \in S_\ell$ as a finite-element function whose nodal values are obtained by averaging the diameters of the triangles within the associated patches:

$$\bar{h}_\ell(a) := \frac{1}{\operatorname{card}(\omega^a_\ell)} \sum_{T \in \omega^a_\ell} h_T, \quad a \in \overline{N}_\ell.$$

The use of the local mesh size makes the heuristic estimates more suitable for adaptively refined meshes. Some of the approximations will employ V_ℓ-realizations of the discrete Lagrangian multipliers σ'_ℓ, $\mu'_\ell \in V_\ell$ defined by

$$(\sigma'_\ell, \varphi^{(a)}_\ell)_{0,\Omega} := \langle\langle \sigma_\ell, \varphi^{(a)}_\ell \rangle\rangle, \quad (\mu'_\ell, \varphi^{(a)}_\ell)_{0,\Omega} := \langle\langle \mu_\ell, \varphi^{(a)}_\ell \rangle\rangle, \ \forall\, a \in \mathcal{N}_\ell.$$

Using the convention of Section 4.4.2, we denote the vectors of coefficients of σ'_ℓ, μ'_ℓ by $\boldsymbol{\sigma}'_\ell$, $\boldsymbol{\mu}'_\ell \in \mathbb{R}^{N_\ell}$ and obtain

$$\boldsymbol{\sigma}_\ell = M_{N_\ell N_\ell} \boldsymbol{\sigma}'_\ell, \quad \boldsymbol{\mu}_\ell = M_{N_\ell N_\ell} \boldsymbol{\mu}'_\ell.$$

Thus, the values of σ'_ℓ, μ'_ℓ in the interior nodes are determined via

$$\boldsymbol{\sigma}'_\ell = \boldsymbol{M}^{-1}_{N_\ell N_\ell} \boldsymbol{\sigma}_\ell, \quad \boldsymbol{\mu}'_\ell = \boldsymbol{M}^{-1}_{N_\ell N_\ell} \boldsymbol{\mu}_\ell.$$

The approximations $\chi_{\ell,\mathcal{A}} \in S_\ell$, $\chi_{\ell,\mathcal{Z}} \in V_\ell$ of the characteristic functions $\chi_{\mathcal{A}}$, $\chi_{\mathcal{Z}}$ are defined by means of

$$\chi_{\ell,\mathcal{A}}(a) := 1 - \frac{(\psi_\ell - y_\ell)(a)}{\gamma \bar{h}^r_\ell(a) + (\psi_\ell - y_\ell)(a)}, \quad a \in \overline{\mathcal{N}}_\ell, \tag{6.6.1a}$$

$$\chi_{\ell,\mathcal{Z}}(a) := 1 - \frac{\sigma'_\ell(a)}{\gamma \bar{h}^r_\ell(a) + \sigma'_\ell(a)}, \quad a \in \mathcal{N}_\ell, \tag{6.6.1b}$$

where $0 < \gamma \leq 1$ and $r > 0$ are fixed. In case of uniform meshes with $\bar{h}_\ell \approx h_\ell := \max_{T \in \mathcal{T}_\ell} h_T$, the following result reflects the approximation properties of $\chi_{\ell,\mathcal{A}}$, $\chi_{\ell,\mathcal{Z}}$ (cf. [47]).

Proposition 6.6.1. *For $\varepsilon \in [0,1)$ and γ, r as in (6.6.1) consider the partition*

$$\mathcal{I} \cap \mathcal{I}_\ell = \mathcal{I}_1 \cup \mathcal{I}_2,$$

where the sets $\mathcal{I}_k$, $1 \le k \le 2$, are given by

$$\begin{aligned} \mathcal{I}_1 &:= \{x \in \mathcal{I} \,|\, 0 < \psi_\ell(x) - y_\ell(x) \le \gamma h_\ell^{\varepsilon r}\}, \\ \mathcal{I}_2 &:= \{x \in \mathcal{I} \,|\, \psi_\ell(x) - y_\ell(x) > \gamma h_\ell^{\varepsilon r}\}. \end{aligned}$$

Then, it holds

$$\|\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}}\|_{0,\omega} \begin{cases} = & 0 & , \ \omega \subset \mathcal{A} \cap \mathcal{A}_\ell \\ < & \min\{|\omega|^{1/2}, \gamma^{-1} h_\ell^{-r} \|y_\ell - \psi_\ell\|_{0,\omega}\} & , \ \omega \subset \mathcal{A} \cap \mathcal{I}_\ell \\ = & |\omega|^{1/2} & , \ \omega \subset \mathcal{I} \cap \mathcal{A}_\ell \\ < & |\omega|^{1/2} & , \ \omega \subset \mathcal{I}_1 \\ < & |\omega|^{1/2} h_\ell^{r(1-\varepsilon)} & , \ \omega \subset \mathcal{I}_2 \end{cases}.$$

Proof. Without loss of generality we may assume $h_\ell \le 1$. For the proof of the first assertion we distinguish several cases.
Case 1 ($\omega \subset \mathcal{A} \cap \mathcal{A}_\ell$): $\chi_{\ell,\mathcal{A}}|_\omega = \chi_{\mathcal{A}}|_\omega = 1$ and thus $\|\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}}\|_{0,\omega} = 0$.
Case 2 ($\omega \subset \mathcal{A} \cap \mathcal{I}_\ell$): We have $\chi_{\mathcal{A}}|_\omega = 1$ and hence,

$$(\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}})|_\omega = \frac{(\psi_\ell - y_\ell)|_\omega}{\gamma h_\ell^r + (\psi_\ell - y_\ell)|_\omega}.$$

Since $(\psi_\ell - y_\ell)|_\omega > 0$ and $\gamma h_\ell^r > 0$, it follows that

$$(\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}})|_\omega < \gamma^{-1} h_\ell^{-r} (\psi_\ell - y_\ell)|_\omega \quad \text{and} \quad (\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}})|_\omega < 1.$$

Thus, $\|\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}}\|_{0,\omega} < \min\{|\omega|^{1/2}, \gamma^{-1} h_\ell^{-r} \|\psi_\ell - y_\ell\|_{0,\omega}\}$.
Case 3 ($\omega \subset \mathcal{I} \cap \mathcal{A}_\ell$): Since $\chi_{\mathcal{A}}|_\omega = 0$ and $\chi_{\ell,\mathcal{A}}|_\omega = 1$, we obtain $\|\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}}\|_{0,\omega} = |\omega|^{1/2}$.
Case 4 ($\omega \subset \mathcal{I} \cap \mathcal{I}_\ell$): We have $\chi_{\mathcal{A}}|_\omega = 0$ and

$$(\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}})|_\omega = \frac{\gamma h_\ell^r}{\gamma h_\ell^r + (\psi_\ell - y_\ell)|_\omega}.$$

For $\omega \subset \mathcal{I}_1$ this implies $(\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}})|_\omega < 1$ and we obtain $\|\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}}\|_{0,\omega} < |\omega|^{1/2}$.
On the other hand, for $\omega \subset \mathcal{I}_2$, taking $h_\ell \le 1$ into account, we find

$$(\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}})|_\omega < \min\{1, h_\ell^{r(1-\varepsilon)}\} = h_\ell^{r(1-\varepsilon)},$$

which leads to $\|\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}}\|_{0,\omega} < |\omega|^{1/2} h_\ell^{r(1-\varepsilon)}$. □

The latter proposition implies

$$\begin{aligned} \|\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}}\|_{0,\Omega}^2 \le \min\{&|\mathcal{A} \cap \mathcal{I}_\ell|, \gamma^{-2} h_\ell^{-2r} (\|y_\ell - y\|_{0,\mathcal{A}\cap\mathcal{I}_\ell} + \|\psi_\ell - \psi\|_{0,\mathcal{A}\cap\mathcal{I}_\ell})^2\} \\ &+ |\mathcal{I} \cap \mathcal{A}_\ell| + |\mathcal{I}_1| + h_\ell^{2r(1-\varepsilon)} |\mathcal{I}_2|. \end{aligned}$$

Herewith, if

$$|\mathcal{A} \cap \mathcal{I}_\ell| \to 0,\ \ell \to \infty \quad \vee \quad \|y_\ell - y\|_{0,\mathcal{A}\cap\mathcal{I}_\ell} + \|\psi_\ell - \psi\|_{0,\mathcal{A}\cap\mathcal{I}_\ell} = O(h_\ell^q),\ q > r, \tag{6.6.2a}$$

$$|\mathcal{I} \cap \mathcal{A}_\ell| \to 0,\ |\mathcal{I}_1| \to 0,\ \ell \to \infty, \tag{6.6.2b}$$

we attain L^2-convergence of $\chi_{\ell,\mathcal{A}}$ to $\chi_{\mathcal{A}}$.

Corollary 6.6.2. *If* (6.6.2) *hold true, we have* $\|\chi_{\mathcal{A}} - \chi_{\ell,\mathcal{A}}\|_{0,\Omega} \to 0,\ \ell \to \infty$.

Similar result can be shown for $\chi_{\ell,\mathcal{Z}}$.

6.6.2 Approximation of the continuous sets

Based on the approximations $\chi_{\ell,\mathcal{A}}$, $\chi_{\ell,\mathcal{Z}}$ of the characteristic functions of the continuous sets $\mathcal{A}$ and $\mathcal{Z}$, we can derive approximations of the continuous active, inactive, strongly active, zero, and biactive sets. To this end, for $0 < \kappa \leq 1$ and $0 < r' \leq r$, based on the nodal values of $\chi_{\ell,\mathcal{A}}$ and $\chi_{\ell,\mathcal{Z}}$, we first define nodal sets $\bar{A}_\ell$, $\bar{I}_\ell$, $\bar{C}_\ell$, $\bar{Z}_\ell$, and $\bar{B}_\ell$ as approximations of their continuous counterparts $\mathcal{A}$, $\mathcal{I}$, $\mathcal{C}$, $\mathcal{Z}$, and $\mathcal{B}$ according to

$$\begin{aligned}
\bar{A}_\ell &:= \{a \in \overline{\mathcal{N}}_\ell \mid \chi_{\ell,\mathcal{A}}(a) \geq 1 - \kappa h_\ell^{r'}\}, & \bar{I}_\ell &:= \overline{\mathcal{N}}_\ell \setminus \bar{A}_\ell,\\
\bar{C}_\ell &:= \Big[\mathcal{N}_\ell \setminus \{a \in \mathcal{N}_\ell \mid \chi_{\ell,\mathcal{Z}}(a) \geq 1 - \kappa h_\ell^{r'}\}\Big] \cap \bar{A}_\ell, & \bar{Z}_\ell &:= \mathcal{N}_\ell \setminus \bar{C}_\ell,\\
\bar{B}_\ell &:= \bar{A}_\ell \cap \bar{Z}_\ell.
\end{aligned}$$

Remark 6.6.3. The definition of the set $\bar{C}_\ell$ guarantees $\bar{C}_\ell \subseteq \bar{A}_\ell$, and, hence, aligns the properties of $\bar{C}_\ell$, $\bar{A}_\ell$ with that of the continuous active and strongly active sets (cf. Proposition 2.4.9). Furthermore, it is easy to see that the definition of $\bar{A}_\ell$ implies $A_\ell \subseteq \bar{A}_\ell$. An analog of this inclusion in the continuous setting can be found in Chapter 5. There we had shown that on a sequence of limit dense meshes, for $\bar{\ell} \in \mathcal{N}$ sufficiently large, there holds $\mathcal{I} \subseteq \mathcal{I}_\ell$, $\ell \geq \bar{\ell}$ (cf. (5.2.8)). Thus, $\mathcal{A}_\ell \subseteq \mathcal{A}$, $\ell \geq \bar{\ell}$.

These sets constitute a suitable basis for the specification of approximations $\bar{\mathcal{A}}_\ell$ of $\mathcal{A}$, $\bar{\mathcal{I}}_\ell$ of $\mathcal{I}$, $\bar{\mathcal{C}}_\ell$ of $\mathcal{C}$, and $\bar{\mathcal{Z}}_\ell$ of $\mathcal{Z}$ by means of

$$\bar{\mathcal{A}}_\ell := \bigcup\{T \in \bar{\mathcal{A}}_\ell^T\}, \quad \bar{\mathcal{A}}_\ell^T := \{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subseteq \bar{A}_\ell\}, \tag{6.6.3a}$$

$$\bar{\mathcal{I}}_\ell := \bigcup\{T \in \bar{\mathcal{I}}_\ell^T \cup \bar{\mathcal{F}}_{y_\ell}^T\}, \quad \bar{\mathcal{I}}_\ell^T := \{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subseteq \bar{I}_\ell\}, \tag{6.6.3b}$$

$$\begin{aligned}
\bar{\mathcal{C}}_\ell := \bigcup\{T \in \bar{\mathcal{C}}_\ell^T\}, \quad \bar{\mathcal{C}}_\ell^T &:= \{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subseteq \bar{C}_\ell\} \cup \\
&\{T \in \mathcal{T}_\ell \setminus \mathcal{T}_\ell(\Omega) \mid \varnothing \neq \mathcal{N}_\ell(T) \cap \mathcal{N}_\ell \subseteq \bar{C}_\ell \ \wedge\ T \subseteq \bar{\mathcal{A}}_\ell^T\},
\end{aligned} \tag{6.6.3c}$$

$$\bar{\mathcal{Z}}_\ell := \bigcup\{T \in \bar{\mathcal{Z}}_\ell^T \cup \bar{\mathcal{F}}_{\sigma_\ell}^T\}, \quad \bar{\mathcal{Z}}_\ell^T := \{T \in \mathcal{T}_\ell \mid \mathcal{N}_\ell(T) \subseteq \bar{Z}_\ell \cup \mathcal{N}_\ell^\Gamma\}. \tag{6.6.3d}$$

The biactive set $\mathcal{B}$ and the free boundaries $\mathcal{F}(y)$, $\mathcal{F}(\sigma)$ are approximated by

$$\bar{\mathcal{B}}_\ell := \bigcup\{T \in \bar{\mathcal{B}}_\ell^T\}, \quad \bar{\mathcal{B}}_\ell^T := \bar{\mathcal{A}}_\ell^T \setminus \bar{\mathcal{C}}_\ell^T, \tag{6.6.3e}$$

$$\bar{\mathcal{F}}_{y_\ell} := \bigcup\{T \in \bar{\mathcal{F}}_{y_\ell}^T\}, \quad \bar{\mathcal{F}}_{y_\ell}^T := \mathcal{T}_\ell \setminus (\bar{\mathcal{A}}_\ell^T \cup \bar{\mathcal{I}}_\ell^T), \tag{6.6.3f}$$

$$\bar{\mathcal{F}}_{\sigma_\ell} := \bigcup\{T \in \bar{\mathcal{F}}_{\sigma_\ell}^T\}, \quad \bar{\mathcal{F}}_{\sigma_\ell}^T := \mathcal{T}_\ell \setminus (\bar{\mathcal{C}}_\ell^T \cup \bar{\mathcal{Z}}_\ell^T). \tag{6.6.3g}$$

Remark 6.6.4. In definition (6.6.3c) we followed the scheme introduced in Section 4.3 by giving the triangles adjacent to the boundary a special treatment. As a result, the approximations $\bar{\mathcal{A}}_\ell$, $\bar{\mathcal{C}}_\ell$, $\bar{\mathcal{B}}_\ell$ have properties alike the discrete sets $\mathcal{A}_\ell$, $\mathcal{C}_\ell$, $\mathcal{B}_\ell$. Namely, $\bar{\mathcal{C}}_\ell \subseteq \bar{\mathcal{A}}_\ell$ and $\bar{\mathcal{B}}_\ell = \varnothing$ as soon as $\bar{B}_\ell = \varnothing$ and $\{T \in \mathcal{T}_\ell \,|\, \mathcal{N}_\ell(T) \subseteq \mathcal{N}_\ell^\Gamma\} = \varnothing$ (cf. Proposition 4.3.14).

The numerical results documented in Chapter 8, will allow us to measure the quality of the approximation of the active set $\mathcal{A}$ and the strongly active set $\mathcal{C}$ by the a posteriori quantities

$$e_{\ell,\mathcal{A}}^{dva} := \|\chi_{\mathcal{A}_\ell} - \chi_{\bar{\mathcal{A}}_\ell}\|_{L^1(\Omega)}, \qquad e_{\ell,\mathcal{C}}^{dva} := \|\chi_{\mathcal{C}_\ell} - \chi_{\bar{\mathcal{C}}_\ell}\|_{L^1(\Omega)}, \tag{6.6.4a}$$

where the upper index "dva" stands for "discrete versus approximate". Furthermore, (6.6.4a) will be compared with the quantities

$$e_{\ell,\mathcal{A}}^{evd} := \|\chi_{\mathcal{A}} - \chi_{\mathcal{A}_\ell}\|_{L^1(\Omega)}, \qquad e_{\ell,\mathcal{C}}^{evd} := \|\chi_{\mathcal{C}} - \chi_{\mathcal{C}_\ell}\|_{L^1(\Omega)}, \tag{6.6.4b}$$

$$e_{\ell,\mathcal{A}}^{eva} := \|\chi_{\mathcal{A}} - \chi_{\bar{\mathcal{A}}_\ell}\|_{L^1(\Omega)}, \qquad e_{\ell,\mathcal{C}}^{eva} := \|\chi_{\mathcal{C}} - \chi_{\bar{\mathcal{C}}_\ell}\|_{L^1(\Omega)}, \tag{6.6.4c}$$

where, the upper indices "evd" and "eva" mean "exact versus discrete" and "exact versus approximate", respectively. Obviously, the latter quantities are only available if the exact solution is known.

Remark 6.6.5. In all the numerical experiments, the values $\kappa = \gamma = 1$ and $r' = r = 1$ were used. These values yield good performance of the approximations $\bar{\mathcal{A}}_\ell$, $\bar{\mathcal{C}}_\ell$ in terms of the quantities (6.6.4). If required, the approximations of the exact sets can be fine-tuned by adjusting the values of the parameters κ, γ, r, and r'.

6.6.3 Approximation of the continuous states and multipliers

In this section we will derive approximations of the state y and the adjoint state p as well as various approximations of the multipliers σ and μ in terms of the approximations of the continuous active, inactive, strongly active, zero, and biactive nodal points (sets) and free boundaries provided in the previous paragraph. Motivated by superconvergence results by local averaging (cf., e.g., [4]), we define approximations $\bar{y}_\ell \in V_\ell$ of y and $\bar{p}_\ell \in V_\ell$ of p according to

$$\bar{y}_\ell(a) := \begin{cases} \operatorname{card}(\mathcal{N}_\ell(\omega_\ell^a))^{-1} \displaystyle\sum_{a' \in \mathcal{N}_\ell(\omega_\ell^a)} y_\ell(a'), & a \in \bar{I}_\ell \\ \psi_\ell(a) & , \ a \in \bar{A}_\ell \end{cases}, \tag{6.6.5a}$$

$$\bar{p}_\ell(a) := \begin{cases} \operatorname{card}(\mathcal{N}_\ell(\omega_\ell^a))^{-1} \displaystyle\sum_{a' \in \mathcal{N}_\ell(\omega_\ell^a)} p_\ell(a'), & a \in \bar{Z}_\ell \\ 0 & , \ a \in \bar{C}_\ell \end{cases}. \tag{6.6.5b}$$

Likewise, we define approximations σ_ℓ'' of σ and μ_ℓ'' of μ by means of

$$\sigma_\ell''(a) := \begin{cases} \operatorname{card}(\mathcal{N}_\ell(\omega_\ell^a))^{-1} \sum\limits_{a' \in \mathcal{N}_\ell(\omega_\ell^a)} \sigma_\ell'(a'), & a \in \bar{C}_\ell \\ 0 & , \ a \in \bar{Z}_\ell \end{cases}, \tag{6.6.5c}$$

$$\mu_\ell''(a) := \begin{cases} \operatorname{card}(\mathcal{N}_\ell(\omega_\ell^a))^{-1} \sum\limits_{a' \in \mathcal{N}_\ell(\omega_\ell^a)} \mu_\ell'(a'), & a \in \bar{I}_\ell \\ 0 & , \ a \in \bar{A}_\ell \end{cases}. \tag{6.6.5d}$$

Remark 6.6.6. The functions $\bar{y}_\ell$, $\bar{p}_\ell$ will replace y, p in the approximation of the consistency error of $e^c_{\ell,rel}$, whereas σ_ℓ'' and μ_ℓ'' will be used in the approximation of $e^c_{\ell,eff}$ and in a special form of the approximations of $e^c_{\ell,rel}$.

For one form of approximation of the consistency error $e^c_{\ell,rel}$, an approximation of the multipliers σ, μ is used which relies the structural properties of the solution of the continuous problem (y, u, σ, p, μ). Suppose that the following assumption on the structure of the sets $\mathcal{A}$ and $\mathcal{C}$ is satisfied.

Assumption 6.6.7. *There holds $\mathcal{C} = \bigcup_{i=1}^s \mathcal{C}_i$ and $\mathcal{A} = \bigcup_{i=1}^q \mathcal{A}_i$ with $\mathcal{C}_i$, $i = 1, \dots, s$ and $\mathcal{A}_i$, $i = 1, \dots, q$ being connected, pairwise disjoint Lipschitz sets.*

Then for any $v \in V$, Proposition 2.3.12 guarantees existence of sets $\widetilde{\mathcal{C}}$, $\widetilde{\mathcal{A}}$ and functions $v_{\mathcal{C}}^{ext}$, $v_{\mathcal{A}}^{ext} \in V$ such that $\mathcal{C} \subseteq \widetilde{\mathcal{C}} \subseteq \Omega$, $\mathcal{A} \subseteq \widetilde{\mathcal{A}} \subseteq \Omega$ and

$$\begin{aligned} \langle \sigma, v \rangle &= \langle \sigma, v_{\mathcal{C}}^{ext} \rangle = (f + u, v_{\mathcal{C}}^{ext})_{0,\widetilde{\mathcal{C}}} - (\nabla y, \nabla v_{\mathcal{C}}^{ext})_{0,\widetilde{\mathcal{C}}}, \\ \langle \mu, v \rangle &= \langle \mu, v_{\mathcal{A}}^{ext} \rangle = (y^d - y, v_{\mathcal{A}}^{ext})_{0,\widetilde{\mathcal{A}}} - (\nabla p, \nabla v_{\mathcal{A}}^{ext})_{0,\widetilde{\mathcal{A}}}. \end{aligned}$$

Employing the structural information provided in Section 3.3.3, we obtain

$$\begin{aligned} \langle \sigma, v \rangle &= \left[(f + u^d, v)_{0,\mathcal{C}} - (\nabla \psi, \nabla v)_{0,\mathcal{C}} \right] - \left[(\Delta \psi, v_{\mathcal{C}}^{ext})_{0,(\widetilde{\mathcal{C}} \setminus \mathcal{C}) \cap \mathcal{B}} + (\nabla \psi, \nabla v_{\mathcal{C}}^{ext})_{0,(\widetilde{\mathcal{C}} \setminus \mathcal{C}) \cap \mathcal{B}} \right] \\ &\quad + \left[(f + u^d + \alpha^{-1} p, v_{\mathcal{C}}^{ext})_{0,(\widetilde{\mathcal{C}} \setminus \mathcal{C}) \cap \mathcal{I}} - (\nabla y, \nabla v_{\mathcal{C}}^{ext})_{0,(\widetilde{\mathcal{C}} \setminus \mathcal{C}) \cap \mathcal{I}} \right], \end{aligned} \tag{6.6.6a}$$

$$\begin{aligned} \langle \mu, v \rangle &= (y^d - \psi, v)_{0,\mathcal{A}} + \alpha (\nabla(\Delta \psi + f + u^d), \nabla v)_{0,\mathcal{B}} \\ &\quad + \left[(y^d - y, v_{\mathcal{A}}^{ext})_{0,(\widetilde{\mathcal{A}} \setminus \mathcal{A})} - (\nabla p, \nabla v_{\mathcal{A}}^{ext})_{0,(\widetilde{\mathcal{A}} \setminus \mathcal{A})} \right]. \end{aligned} \tag{6.6.6b}$$

Remark 6.6.8. Assumption 6.6.7 can be omitted if instead of $\mathcal{C}$, $\mathcal{A}$ we apply Proposition 2.3.12 to unions of finite numbers of minimal disjoint Lipschitz sets $\widetilde{Lip}(\mathcal{C})$, $\widetilde{Lip}(\mathcal{A})$ such that $\mathcal{C} \subseteq \widetilde{Lip}(\mathcal{C})$, $\mathcal{A} \subseteq \widetilde{Lip}(\mathcal{A})$. In this case, sets $\widetilde{\mathcal{C}}$ and $\widetilde{\mathcal{A}}$ resulting from Proposition 2.3.12 will satisfy $\mathcal{C} \subseteq \widetilde{Lip}(\mathcal{C}) \subseteq \widetilde{\mathcal{C}} \subseteq \Omega$, $\mathcal{A} \subseteq \widetilde{Lip}(\mathcal{A}) \subseteq \widetilde{\mathcal{A}} \subseteq \Omega$, whereas the representation of the exact multipliers (6.6.6) remains unchanged.

In order to provide a fully computable approximation, we replace the unknown sets $\mathcal{C}$, $\mathcal{B}$, $\mathcal{A}$, $\mathcal{I}$ and the unknown functions y, p by their previously defined approximations $\bar{\mathcal{C}}_\ell$, $\bar{\mathcal{B}}_\ell$, $\bar{\mathcal{A}}_\ell$, $\bar{\mathcal{I}}_\ell$, and $\bar{y}_\ell$, $\bar{p}_\ell$. Moreover, $\widetilde{\mathcal{C}}$, $\widetilde{\mathcal{A}}$ are chosen according to

$$\widetilde{\mathcal{C}} := \bar{\mathcal{C}}_\ell \cap \bar{\mathcal{F}}_{\sigma_\ell}, \quad \widetilde{\mathcal{A}} := \bar{\mathcal{A}}_\ell \cap \bar{\mathcal{F}}_{y_\ell},$$

whereas $v_{\mathcal{C}}^{ext}$ and $v_{\mathcal{A}}^{ext}$ are approximated by

$$v_{\bar{\mathcal{C}}_\ell}^{ext} := I_{\bar{C}_\ell}(v_\ell), \quad v_{\bar{\mathcal{A}}_\ell}^{ext} := I_{\bar{A}_\ell}(v_\ell), \quad \forall v_\ell \in V_\ell.$$

Here, I_{D_ℓ}, $D_\ell \subseteq \mathcal{N}_\ell$ is the operator from Definition 4.5.1.

Remark 6.6.9. Obviously, $\operatorname{supp}(v_{\bar{C}_\ell}^{ext}) = \widetilde{\mathcal{C}}$, $\operatorname{supp}(v_{\bar{A}_\ell}^{ext}) = \widetilde{A}$. Hence, this choice of $v_{\bar{\mathcal{C}}_\ell}^{ext}$, $\widetilde{\mathcal{C}}$ and $v_{\bar{\mathcal{A}}_\ell}^{ext}$, $\widetilde{\mathcal{A}}$ aligns with Proposition 2.3.12.

Using the previous approximations in (6.6.6) and assuming sufficient regularity of the data in $\bar{\mathcal{B}}_\ell$, we obtain the following approximations of the action of σ, μ on functions in V_ℓ:

$$\begin{aligned}
\langle \sigma, v_\ell \rangle \approx \langle \bar{\sigma}_\ell^{(1)}, v_{\bar{\mathcal{C}}_\ell}^{ext} \rangle = &\sum_{T \subseteq \bar{\mathcal{C}}_\ell} \Big[(f + u^d, v_\ell)_{0,T} - (\nabla\psi, \nabla v_\ell)_{0,T} \Big] \qquad (6.6.7a)\\
&- \sum_{T \subseteq \bar{\mathcal{F}}_{\sigma_\ell} \cap \bar{\mathcal{B}}_\ell} \Big[(\Delta\psi, I_{\bar{C}_\ell}(v_\ell))_{0,T} + (\nabla\psi, \nabla I_{\bar{C}_\ell}(v_\ell))_{0,T} \Big]\\
&+ \sum_{T \subseteq \bar{\mathcal{F}}_{\sigma_\ell} \cap \bar{\mathcal{I}}_\ell} \Big[(f + u^d + \alpha^{-1}\bar{p}_\ell, I_{\bar{C}_\ell}(v_\ell))_{0,T} - (\nabla \bar{y}_\ell, \nabla I_{\bar{C}_\ell}(v_\ell))_{0,T} \Big],\\
\langle \mu, v_\ell \rangle \approx \langle \bar{\mu}_\ell^{(1)}, v_{\bar{\mathcal{A}}_\ell}^{ext} \rangle = &\sum_{T \subseteq \bar{\mathcal{A}}_\ell} (y^d - \psi, v_\ell)_{0,T} + \sum_{T \subseteq \bar{\mathcal{B}}_\ell} \alpha (\nabla(\Delta\psi + f + u^d), \nabla v_\ell)_{0,T} \qquad (6.6.7b)\\
&+ \sum_{T \subseteq \bar{\mathcal{F}}_{y_\ell}} \Big[(y^d - \bar{y}_\ell, I_{\bar{A}_\ell}(v_\ell))_{0,T} - (\nabla \bar{p}_\ell, \nabla I_{\bar{A}_\ell}(v_\ell))_{0,T} \Big].
\end{aligned}$$

As far as the regularity of the data is concerned, (6.6.7) are well-defined if there holds

$$\Delta\psi \in L^2(\bar{\mathcal{B}}_\ell), \quad \Delta\psi + f + u^d \in H^1(\bar{\mathcal{B}}_\ell). \qquad (6.6.8)$$

Remark 6.6.10. In paragraph 3.3.3, we have seen that $\Delta\psi \in L^2(\mathcal{B})$ and $\Delta\psi + f + u^d \in H^1(\mathcal{B})$, hence, condition (6.6.8) is automatically satisfied if $\bar{\mathcal{B}}_\ell \subset \mathcal{B}$. However, as we do not know $\mathcal{B}$, this criterion is useless in practice, whereas (6.6.8) itself can often be verified.

For the cases where the data of the problem disagree with (6.6.8), the following simplification of (6.6.7) is suggested, where $\bar{y}_\ell$, $\bar{p}_\ell$ replace the exact expressions of y and p in $\bar{\mathcal{B}}_\ell$:

$$\begin{aligned}
\langle \bar{\sigma}_\ell^{(2)}, v_\ell \rangle = &\sum_{T \subseteq \bar{\mathcal{C}}_\ell} \Big[(f + u^d, v_\ell)_{0,T} - (\nabla\psi, \nabla v_\ell)_{0,T} \Big] - \sum_{T \subseteq \bar{\mathcal{F}}_{\sigma_\ell} \cap \bar{\mathcal{B}}_\ell} (\nabla\psi, \nabla I_{\bar{C}_\ell}(v_\ell))_{0,T}\\
&+ \sum_{T \subseteq \bar{\mathcal{F}}_{\sigma_\ell} \cap \bar{\mathcal{I}}_\ell} \Big[(f + u^d + \alpha^{-1}\bar{p}_\ell, I_{\bar{C}_\ell}(v_\ell))_{0,T} - (\nabla \bar{y}_\ell, \nabla I_{\bar{C}_\ell}(v_\ell))_{0,T} \Big], \qquad (6.6.9a)\\
\langle \bar{\mu}_\ell^{(2)}, v_\ell \rangle = &\sum_{T \subseteq \bar{\mathcal{A}}_\ell} (y^d - \psi, v_\ell)_{0,T} - \sum_{T \in \bar{\mathcal{B}}_\ell^T} (\nabla \bar{p}_\ell, \nabla v_\ell)_{0,T}\\
&+ \sum_{T \subseteq \bar{\mathcal{F}}_{y_\ell}} \Big[(y^d - \bar{y}_\ell, I_{\bar{A}_\ell}(v_\ell))_{0,T} - (\nabla \bar{p}_\ell, \nabla I_{\bar{A}_\ell}(v_\ell))_{0,T} \Big]. \qquad (6.6.9b)
\end{aligned}$$

6.6.4 Computable counterparts of the consistency errors

For the consistency error $e^c_{\ell,rel}$, three different types of approximations will be used:

$$e^c_{\ell,rel} \approx E^{c,k}_{\ell,rel} := \bar{e}^{(1,k)}_{\ell,\sigma} + \bar{e}^{(2,k)}_{\ell,\sigma} + \bar{e}^{(1,k)}_{\ell,\mu} + \bar{e}^{(2,k)}_{\ell,\mu}, \quad k = \{1,2,3\}.$$

For the first two approximations $\bar{e}^{c,k}_{\ell,rel}$, $k = \{1,2\}$, we will use the approximation of the multipliers by (6.6.7) and (6.6.9):

$$\begin{aligned} \bar{e}^{(1,k)}_{\ell,\sigma} &:= \langle \tilde{\sigma}_\ell - \bar{\sigma}^{(k)}_\ell, \bar{y}_\ell - y_\ell \rangle, \quad \bar{e}^{(2,k)}_{\ell,\sigma} := \langle \tilde{\sigma}_\ell - \bar{\sigma}^{(k)}_\ell, p_\ell - \bar{p}_\ell \rangle, \\ \bar{e}^{(1,k)}_{\ell,\mu} &:= \langle \tilde{\mu}_\ell - \bar{\mu}^{(k)}_\ell, \bar{y}_\ell - y_\ell \rangle, \quad \bar{e}^{(2,k)}_{\ell,\mu} := \langle \tilde{\mu}_\ell - \bar{\mu}^{(k)}_\ell, \bar{p}_\ell - p_\ell \rangle. \end{aligned}$$

The third approximation is obtained by using the approximation of the multipliers by local averaging (cf. (6.6.5c), (6.6.5d)):

$$\begin{aligned} \bar{e}^{(1,3)}_{\ell,\sigma} &:= \langle \tilde{\sigma}_\ell - \sigma''_\ell, \bar{y}_\ell - y_\ell \rangle, \quad \bar{e}^{(2,3)}_{\ell,\sigma} := \langle \tilde{\sigma}_\ell - \sigma''_\ell, p_\ell - \bar{p}_\ell \rangle, \\ \bar{e}^{(1,3)}_{\ell,\mu} &:= \langle \tilde{\mu}_\ell - \mu''_\ell, \bar{y}_\ell - y_\ell \rangle, \quad \bar{e}^{(2,3)}_{\ell,\mu} := \langle \tilde{\mu}_\ell - \mu''_\ell, \bar{p}_\ell - p_\ell \rangle. \end{aligned}$$

In all three types of approximations, it is convenient to employ the representation (4.5.2) of the extensions $\tilde{\sigma}_\ell$, $\tilde{\mu}_\ell$.

For the approximation of the consistency error $e^c_{\ell,eff}$ we will use the approximation of the multipliers by local averaging as given by (6.6.5c) and (6.6.5d).

$$e^c_{\ell,eff} \lesssim e^{c,L^2}_{\ell,eff} \approx E^c_{\ell,eff} := \sum_{T \subseteq \mathcal{Z}_\ell} h_T^2 \|\sigma''_\ell\|^2_{0,T} + \sum_{T \subseteq \mathcal{I}_\ell} h_T^2 \|\mu''_\ell\|^2_{0,T}. \tag{6.6.10}$$

Definition 6.6.11. *In the following, it will be referred to*

(*i*) $E^{c,k}_{\ell,rel}$ *as the heuristic estimates of* $e^c_{\ell,rel}$ *of type* k, $k = \{1,2,3\}$;

(*ii*) $E^c_{\ell,eff}$ *as the heuristic estimate of* $e^c_{\ell,eff}$.

Once again, it should be emphasized that $E^{c,1}_{\ell,rel}$ can only be used if the data of the problem is sufficiently smooth (cf. Remark 6.6.10).

7 Numerical implementation

In this chapter, an adaptive algorithm is introduced, which is based on the a posteriori error estimate (6.1.1). Furthermore, separate steps of this algorithm are addressed in detail.

7.1 Adaptive algorithm

In the theory of adaptive finite element methods, it is conventional to speak of a general adaptive loop in the form

$$\text{SOLVE} \;\rightarrow\; \text{ESTIMATE} \;\rightarrow\; \text{MARK} \;\rightarrow\; \text{REFINE}\,. \qquad (*)$$

Step SOLVE delivers a numerical solution to the discrete problem on a fixed mesh, whereas the next three modules - ESTIMATE, MARK, and REFINE, - are the basic steps of an adaptive method: in step ESTIMATE, an a posteriori error estimate is computed; according to the information provided by the error estimate (element-wise contributions of the estimator η_ℓ and osc_ℓ), step MARK selects a set of elements in the current triangulation that will be refined within step REFINE in order to obtain the next mesh. In practice, let us reshape the loop $(*)$ to obtain the following adaptive algorithm.

Initialization:		set $\ell = 0$, choose $\mathcal{T}_\ell$, SOLVE $\rightarrow$ ESTIMATE on $\mathcal{T}_\ell$, check stopping criterion;
Iteration:	(i)	set $\ell = \ell + 1$, MARK $\rightarrow$ REFINE ;
	(ii)	SOLVE $\rightarrow$ ESTIMATE on $\mathcal{T}_\ell$;
	(iii)	check stopping criterion.

Algorithm 7.1: The adaptive algorithm.

In the stopping criterion, one can, for instance, compare the magnitude of the a posteriori error estimator with an afore specified tolerance TOL. The algorithm is terminated as soon as the condition $\eta_\ell + osc_\ell \leq TOL$ is satisfied. In the numerical experiments, $\mathcal{T}_0$ is chosen such that $\{T \in \mathcal{T}_0 \,|\, \mathcal{N}_\ell(T) \subseteq \mathcal{N}_\ell^\Gamma\} = \varnothing$ (cf. Proposition 4.3.14).

7.2 Numerical solution

In this section, the algorithm is presented that is used for finding stationary points of (4.2.4). Employing the notation of Section 4.4.2, we recall that $(\boldsymbol{y}_\ell, \boldsymbol{\sigma}_\ell, \boldsymbol{u}_\ell, \boldsymbol{p}_\ell, \boldsymbol{\mu}_\ell) \in \mathbb{R}^{N_\ell} \times \mathbb{R}^{N_\ell} \times \mathbb{R}^{\overline{N}_\ell} \times \mathbb{R}^{N_\ell} \times \mathbb{R}^{N_\ell}$ is an S-stationary point if it satisfies

$$\boldsymbol{K}_{N_\ell N_\ell} \boldsymbol{y}_\ell - \alpha^{-1} \boldsymbol{M}_{N_\ell N_\ell} \boldsymbol{p}_\ell = \boldsymbol{M}_{N_\ell \overline{N}_\ell} (\boldsymbol{u}_\ell^d + \boldsymbol{f}_\ell) - \boldsymbol{\sigma}_\ell, \tag{7.2.1a}$$

$$\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell \geq \boldsymbol{0}_{N_\ell}, \quad \boldsymbol{\sigma}_\ell \geq \boldsymbol{0}_{N_\ell}, \quad \boldsymbol{\sigma}_\ell \cdot (\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell) = 0, \tag{7.2.1b}$$

$$\boldsymbol{K}_{N_\ell N_\ell} \boldsymbol{p}_\ell + \boldsymbol{M}_{N_\ell N_\ell} \boldsymbol{y}_\ell = \boldsymbol{M}_{N_\ell \overline{N}_\ell} \boldsymbol{y}_\ell^d - \boldsymbol{\mu}_\ell, \tag{7.2.1c}$$

$$\boldsymbol{u}_\ell(\boldsymbol{n}_\ell) = \alpha^{-1} \boldsymbol{p}_\ell + \boldsymbol{u}_\ell^d(\boldsymbol{n}_\ell), \quad \boldsymbol{u}_\ell(\boldsymbol{n}_\ell^\Gamma) = \boldsymbol{u}_\ell^d(\boldsymbol{n}_\ell^\Gamma), \tag{7.2.1d}$$

$$\boldsymbol{p}_\ell(\boldsymbol{n}_\ell^C) = \boldsymbol{0}_{N_\ell^C}, \tag{7.2.1e}$$

$$\boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell^I) = \boldsymbol{0}_{N_\ell^I}, \tag{7.2.1f}$$

$$\boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell^B) \geq \boldsymbol{0}_{N_\ell^B}, \ \boldsymbol{p}_\ell(\boldsymbol{n}_\ell^B) \geq \boldsymbol{0}_{N_\ell^B}, \tag{7.2.1g}$$

a C-stationary point if it satisfies (7.2.1a) - (7.2.1f) with

$$\boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell^B) * \boldsymbol{p}_\ell(\boldsymbol{n}_\ell^B) \geq \boldsymbol{0}_{N_\ell^B}, \tag{7.2.2}$$

and a stationary point with strict complementarity if it satisfies (7.2.1a) - (7.2.1d) and

$$\boldsymbol{p}_\ell(\boldsymbol{n}_\ell^A) = \boldsymbol{0}_{N_\ell^A}, \quad \boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell^I) = \boldsymbol{0}_{N_\ell^I}. \tag{7.2.3}$$

7.2.1 Algorithm

In step SOLVE of the adaptive algorithm 7.1, stationary points of (4.2.4) on a fixed mesh $\mathcal{T}_\ell$ are computed. For problems that have stationary points with strict complementarity, the so-called primal-dual active set strategy (PDASS) as suggested in [41] is employed. This method is based on an equivalent formulation of the complementarity condition (7.2.1b):

$$\boldsymbol{\sigma}_\ell = \max\left(0, \boldsymbol{\sigma}_\ell + c\,[\boldsymbol{y}_\ell - \boldsymbol{\psi}_\ell]\right), \quad \forall c > 0. \tag{7.2.4}$$

The equivalence of (7.2.1b) to (7.2.4) is obvious if we compare the components of the

	$i \in \boldsymbol{n}_\ell^C$	$i \in \boldsymbol{n}_\ell^B$	$i \in \boldsymbol{n}_\ell^I$
$B_\ell \neq \varnothing$	$(\boldsymbol{y}_\ell - \boldsymbol{\psi}_\ell)_i = 0$	$(\boldsymbol{y}_\ell - \boldsymbol{\psi}_\ell)_i = 0$	$(\boldsymbol{y}_\ell - \boldsymbol{\psi}_\ell)_i < 0$
	$(\boldsymbol{\sigma}_\ell)_i > 0$	$(\boldsymbol{\sigma}_\ell)_i = 0$	$(\boldsymbol{\sigma}_\ell)_i = 0$
$B_\ell = \varnothing \quad (C_\ell = A_\ell)$	$(\boldsymbol{y}_\ell - \boldsymbol{\psi}_\ell)_i = 0$	___	$(\boldsymbol{y}_\ell - \boldsymbol{\psi}_\ell)_i < 0$
	$(\boldsymbol{\sigma}_\ell)_i > 0$		$(\boldsymbol{\sigma}_\ell)_i = 0$

Table 7.2: Components of vectors $\boldsymbol{y}_\ell - \boldsymbol{\psi}_\ell$, $\boldsymbol{\sigma}_\ell$ with indices from $\boldsymbol{n}_\ell^C$, $\boldsymbol{n}_\ell^B$, $\boldsymbol{n}_\ell^I$

vectors $\boldsymbol{\psi}_\ell - \boldsymbol{y}_\ell$ and $\boldsymbol{\sigma}_\ell$ with the indices $\boldsymbol{n}_\ell^C$, $\boldsymbol{n}_\ell^B$, $\boldsymbol{n}_\ell^I$ summarized in Table 7.2 for the cases of strict complementarity ($B_\ell = \varnothing$) and lack of strict complementarity ($B_\ell \neq \varnothing$).

In Algorithm 7.3, every object associated with iteration k is marked by either a subscript (k) or an affix $[k]$.

Initialization:		set $k := 0$ and choose $\boldsymbol{y}_\ell^{(k)}$, $\boldsymbol{\sigma}_\ell^{(k)} \in \mathbb{R}^{N_\ell}$, $A_\ell^{(k)}$, and $c > 0$;
Iteration:	(i)	set $k := k+1$, $\hat{\boldsymbol{\sigma}}_\ell := \max\left(0, \boldsymbol{\sigma}_\ell^{(k-1)} + c\left[\boldsymbol{y}_\ell^{(k-1)} - \boldsymbol{\psi}_\ell\right]\right)$, $A_\ell^{(k)} := \{a_i \in \mathcal{N}_\ell \mid (\hat{\boldsymbol{\sigma}}_\ell)_i > 0\}$, $I_\ell^{(k)} = \mathcal{N}_\ell \setminus A_\ell^{(k)}$;
	(ii)	if $A_\ell^{(k-1)} = A_\ell^{(k)}$, stop;
	(iii)	find $\boldsymbol{y}_\ell^{(k)}$, $\boldsymbol{\sigma}_\ell^{(k)}$, $\boldsymbol{p}_\ell^{(k)}$, $\boldsymbol{\mu}_\ell^{(k)} \in \mathbb{R}^{\overline{N}_\ell}$ such that (7.2.1a) and (7.2.1c) are satisfied simultaneously with $\boldsymbol{\sigma}_\ell^{(k)}(\boldsymbol{n}_\ell^I[k]) = \mathbf{0}_{N_\ell^I[k]}$, $\boldsymbol{y}_\ell^{(k)}(\boldsymbol{n}_\ell^A[k]) = \boldsymbol{\psi}_\ell(\boldsymbol{n}_\ell^A[k])$, $\boldsymbol{\mu}_\ell^{(k)}(\boldsymbol{n}_\ell^I[k]) = \mathbf{0}_{N_\ell^I[k]}$, $\boldsymbol{p}_\ell^{(k)}(\boldsymbol{n}_\ell^A[k]) = \mathbf{0}_{N_\ell^A[k]}$.

Algorithm 7.3: PDASS algorithm.

Let us take a closer look at step (i) of this algorithm. Here, an approximation $A_\ell^{(k)}$ of the unknown set of active nodes A_ℓ is computed using the solution from the previous iteration, $\boldsymbol{\sigma}_\ell^{(k-1)}$, $\boldsymbol{y}_\ell^{(k-1)}$, and the complementarity constraint in form (7.2.4) under the assumption of strict complementarity. Indeed, $\hat{\boldsymbol{\sigma}}_\ell$ constitutes an approximation to σ_ℓ, thus, condition $\{a_i \in \mathcal{N}_\ell \mid (\hat{\boldsymbol{\sigma}}_\ell)_i > 0\}$ of step (i) is a natural characterization of the approximation $C_\ell^{(k)}$ of the strongly active nodes on iteration k. Hence, step (i) implicitly combines two statements:

$$
\begin{aligned}
C_\ell^{(k)} &:= \{a_i \in \mathcal{N}_\ell \mid (\hat{\boldsymbol{\sigma}}_\ell)_i > 0\},\\
A_\ell^{(k)} &:= C_\ell^{(k)}.
\end{aligned}
$$

Leaving assumption $A_\ell^{(k)} := C_\ell^{(k)}$ aside, a more general version of PDASS algorithm can be obtained, later on referred to as PDASS(LSC) (LSC stands for lack of strict complementarity) (see Algorithm 7.4). The steps of PDASS(LSC) are identical to that of PDASS, only the sets $A_\ell^{(k)}$, $I_\ell^{(k)}$ are replaced with $C_\ell^{(k)}$, $Z_\ell^{(k)}$, respectively. Let us now address the steps of Algorithm 7.4 in detail.

Initialization: set $k := 0$ and choose $\boldsymbol{y}_\ell^{(k)}$, $\boldsymbol{\sigma}_\ell^{(k)} \in \mathbb{R}^{N_\ell}$, $C_\ell^{(k)}$, and $c > 0$;

Iteration:

(i) set $k := k+1$, $\hat{\boldsymbol{\sigma}}_\ell := \max\left(0, \boldsymbol{\sigma}_\ell^{(k-1)} + c\left[\boldsymbol{y}_\ell^{(k-1)} - \boldsymbol{\psi}_\ell\right]\right)$,
$C_\ell^{(k)} := \{a_i \in \mathcal{N}_\ell \mid (\hat{\boldsymbol{\sigma}}_\ell)_i > 0\}$, $Z_\ell^{(k)} = \mathcal{N}_\ell \setminus C_\ell^{(k)}$;

(ii) if $C_\ell^{(k-1)} = C_\ell^{(k)}$, stop;

(iii) find $\boldsymbol{y}_\ell^{(k)}$, $\boldsymbol{\sigma}_\ell^{(k)}$, $\boldsymbol{p}_\ell^{(k)}$, $\boldsymbol{\mu}_\ell^{(k)} \in \mathbb{R}^{N_\ell}$ such that (7.2.1a) and (7.2.1c) are satisfied simultaneously with

$$\boldsymbol{\sigma}_\ell^{(k)}(\boldsymbol{n}_\ell^Z[k]) = \boldsymbol{0}_{N_\ell^Z[k]}, \quad \boldsymbol{y}_\ell^{(k)}(\boldsymbol{n}_\ell^C[k]) = \boldsymbol{\psi}_\ell(\boldsymbol{n}_\ell^C[k]),$$
$$\boldsymbol{\mu}_\ell^{(k)}(\boldsymbol{n}_\ell^Z[k]) = \boldsymbol{0}_{N_\ell^Z[k]}, \quad \boldsymbol{p}_\ell^{(k)}(\boldsymbol{n}_\ell^C[k]) = \boldsymbol{0}_{N_\ell^C[k]}$$

Algorithm 7.4: PDASS(LSC) algorithm.

Steps of PDASS(LSC)

Initialization. In the numerical experiments, the algorithm is initialized by solving the unconstrained optimal control problem

$$\begin{aligned} &\textit{minimize} && J_\ell(\boldsymbol{y}_\ell, \boldsymbol{u}_\ell) \\ &\textit{over} && (\boldsymbol{y}_\ell, \boldsymbol{u}_\ell) \in \mathbb{R}^{N_\ell} \times \mathbb{R}^{\overline{N}_\ell}, \\ &\textit{subject to} && \boldsymbol{K}_{N_\ell N_\ell} \boldsymbol{y}_\ell = \boldsymbol{M}_{N_\ell \overline{N}_\ell}(\boldsymbol{u}_\ell + \boldsymbol{f}_\ell). \end{aligned} \tag{7.2.5}$$

Along with the adjoint state $\bar{\boldsymbol{p}}_\ell \in \mathbb{R}^{N_\ell}$, the solution $(\bar{\boldsymbol{y}}_\ell, \bar{\boldsymbol{u}}_\ell) \in \mathbb{R}^{N_\ell} \times \mathbb{R}^{\overline{N}_\ell}$ of the unconstrained problem (7.2.5) satisfies the system

$$\begin{aligned} &\boldsymbol{K}_{N_\ell N_\ell} \bar{\boldsymbol{y}}_\ell - \alpha^{-1} \boldsymbol{M}_{N_\ell N_\ell} \bar{\boldsymbol{p}}_\ell = \boldsymbol{M}_{N_\ell \overline{N}_\ell}(\boldsymbol{u}_\ell^d + \boldsymbol{f}_\ell), \\ &\boldsymbol{K}_{N_\ell N_\ell} \bar{\boldsymbol{p}}_\ell + \boldsymbol{M}_{N_\ell N_\ell} \bar{\boldsymbol{y}}_\ell = \boldsymbol{M}_{N_\ell \overline{N}_\ell} \boldsymbol{y}_\ell^d, \\ &\bar{\boldsymbol{u}}_\ell(\boldsymbol{n}_\ell) = \alpha^{-1} \bar{\boldsymbol{p}}_\ell + \boldsymbol{u}_\ell^d(\boldsymbol{n}_\ell), \quad \bar{\boldsymbol{u}}_\ell(\boldsymbol{n}_\ell^\Gamma) = \boldsymbol{u}_\ell^d(\boldsymbol{n}_\ell^\Gamma). \end{aligned}$$

The initial conditions are set to $\boldsymbol{y}_\ell^{(0)} := \bar{\boldsymbol{y}}_\ell$, $\boldsymbol{\sigma}_\ell^{(0)} := \boldsymbol{0}_{N_\ell}$, and $C_\ell^{(0)} = \varnothing$.
Iteration. In step (i), an approximation $C_\ell^{(k)}$ of the set of strongly active nodal points C_ℓ is computed. Note that in all numerical experiments the value $c = 1$ is used. The algorithm terminates as soon as the stopping criterion $C_\ell^{(k-1)} = C_\ell^{(k)}$ of step (ii) is satisfied. In step (iii), a transformation of the coupled system (7.2.1a), (7.2.1c) in view of the values prescribed to the vectors $\boldsymbol{y}_\ell^{(k)}$, $\boldsymbol{\sigma}_\ell^{(k)}$, $\boldsymbol{p}_\ell^{(k)}$, $\boldsymbol{\mu}_\ell^{(k)}$ in the indices $\boldsymbol{n}_\ell^C[k]$, $\boldsymbol{n}_\ell^Z[k]$ leads to the system

$$\boldsymbol{A}_\ell^{(k)} \boldsymbol{x}_\ell^{(k)} = \boldsymbol{f}_\ell^{(k)},$$

where the vector of unknowns, $\boldsymbol{x}_\ell^{(k)} \in \mathbb{R}^{2N_\ell^Z[k]}$, the matrix $\boldsymbol{A}_\ell^{(k)} \in \mathbb{R}^{2N_\ell^Z[k]\times 2N_\ell^Z[k]}$, and the right-hand side vector $\boldsymbol{f}_\ell^{(k)} \in \mathbb{R}^{2N_\ell^Z[k]}$ are given by

$$\begin{aligned}
\boldsymbol{x}_\ell^{(k)} &:= (\boldsymbol{y}_\ell^{(k)}(\boldsymbol{n}_\ell^Z[k]), \boldsymbol{p}_\ell^{(k)}(\boldsymbol{n}_\ell^Z[k]))^T,\\
\boldsymbol{A}_\ell^{(k)} &:= \begin{bmatrix} \boldsymbol{K}_\ell(\boldsymbol{n}_\ell^Z[k], \boldsymbol{n}_\ell^Z[k]) & -\alpha^{-1}\boldsymbol{M}_\ell(\boldsymbol{n}_\ell^Z[k], \boldsymbol{n}_\ell^Z[k]) \\ \boldsymbol{M}_\ell(\boldsymbol{n}_\ell^Z[k], \boldsymbol{n}_\ell^Z[k]) & \boldsymbol{K}_\ell(\boldsymbol{n}_\ell^Z[k], \boldsymbol{n}_\ell^Z[k]) \end{bmatrix},\\
\boldsymbol{f}_\ell^{(k)} &:= \begin{bmatrix} \boldsymbol{f}_\ell(\boldsymbol{n}_\ell^Z[k]) + \boldsymbol{u}_\ell^d(\boldsymbol{n}_\ell^Z[k]) - \boldsymbol{K}_\ell(\boldsymbol{n}_\ell^Z[k], \boldsymbol{n}_\ell^C[k])\,\boldsymbol{\psi}_\ell(\boldsymbol{n}_\ell^C[k]) \\ \boldsymbol{y}_\ell^{\boldsymbol{d}}(\boldsymbol{n}_\ell^Z[k]) - \boldsymbol{M}_\ell(\boldsymbol{n}_\ell^Z[k], \boldsymbol{n}_\ell^C[k])\,\boldsymbol{\psi}_\ell(\boldsymbol{n}_\ell^C[k]) \end{bmatrix}.
\end{aligned}$$

Note that $\boldsymbol{A}_\ell^{(k)}$ is non-singular as a principle sub-matrix of a non-singular matrix. The unknown parts of $\boldsymbol{\sigma}_\ell^{(k)}$ and $\boldsymbol{\mu}_\ell^{(k)}$, as well as the values of $\boldsymbol{u}_\ell^{(k)}$ associated with the interior nodal points are computed as follows

$$\begin{aligned}
\boldsymbol{\sigma}_\ell^{(k)}(\boldsymbol{n}_\ell^A[k]) &= \boldsymbol{f}_\ell(\boldsymbol{n}_\ell^A[k]) + \boldsymbol{u}_\ell^d(\boldsymbol{n}_\ell^A[k]) - \boldsymbol{K}_\ell(\boldsymbol{n}_\ell^A[k], \boldsymbol{n}_\ell)\,\boldsymbol{y}_\ell^{(k)} + \alpha^{-1}\boldsymbol{M}_\ell(\boldsymbol{n}_\ell^A[k], \boldsymbol{n}_\ell)\,\boldsymbol{p}_\ell^{(k)},\\
\boldsymbol{\mu}_\ell^{(k)}(\boldsymbol{n}_\ell^A[k]) &= \boldsymbol{y}_\ell^{\boldsymbol{d}}(\boldsymbol{n}_\ell^A[k]) - \boldsymbol{K}_\ell(\boldsymbol{n}_\ell^A[k], \boldsymbol{n}_\ell)\,\boldsymbol{p}_\ell^{(k)} - \boldsymbol{M}_\ell(\boldsymbol{n}_\ell^A[k], \boldsymbol{n}_\ell)\,\boldsymbol{y}_\ell^{(k)},\\
\boldsymbol{u}_\ell^{(k)}(\boldsymbol{n}_\ell) &= \alpha^{-1}\boldsymbol{p}_\ell^{(k)} + \boldsymbol{u}_\ell^d(\boldsymbol{n}_\ell).
\end{aligned}$$

Properties of the solution delivered by PDASS(LSC)

Algorithms PDASS(LSC) and PDASS deliver identical iterarions

$$(\boldsymbol{y}_\ell^{(k)}, \boldsymbol{\sigma}_\ell^{(k)}, \boldsymbol{u}_\ell^{(k)}, \boldsymbol{p}_\ell^{(k)}, \boldsymbol{\mu}_\ell^{(k)}), \quad k \geq 0.$$

The only difference between the algorithms amounts to the name used for the set $\{a_i \in \mathcal{N}_\ell \mid (\hat{\boldsymbol{\sigma}}_\ell)_i > 0\}$. In PDASS(LSC) it is denoted by $C_\ell^{(k)}$, in (PDASS) by $A_\ell^{(k)}$. If terminated at iteration $k = k' + 1$, both algorithms deliver the solution

$$(\boldsymbol{y}_\ell, \boldsymbol{\sigma}_\ell, \boldsymbol{u}_\ell, \boldsymbol{p}_\ell, \boldsymbol{\mu}_\ell) := (\boldsymbol{y}_\ell^{(k')}, \boldsymbol{\sigma}_\ell^{(k')}, \boldsymbol{u}_\ell^{(k')}, \boldsymbol{p}_\ell^{(k')}, \boldsymbol{\mu}_\ell^{(k')}).$$

According to [41], point $(\boldsymbol{y}_\ell, \boldsymbol{\sigma}_\ell, \boldsymbol{u}_\ell, \boldsymbol{p}_\ell, \boldsymbol{\mu}_\ell)$ solves the system (7.2.1a) - (7.2.1d), (7.2.3) for $A_\ell := A_\ell^{(k')}$, i.e., for the discrete optimal control problem (4.2.4) it is a stationary point with strict complementarity. Thus, setting $C_\ell := C_\ell^{(k')} = A_\ell^{(k')}$, the latter observation implies that point $(\boldsymbol{y}_\ell, \boldsymbol{\sigma}_\ell, \boldsymbol{u}_\ell, \boldsymbol{p}_\ell, \boldsymbol{\mu}_\ell)$, viewed as a solution of PDASS(LSC), satisifies the system (7.2.1a) - (7.2.1d) with

$$\boldsymbol{p}_\ell(\boldsymbol{n}_\ell^C) = \boldsymbol{0}_{N_\ell^C}, \quad \boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell^Z) = \boldsymbol{0}_{N_\ell^Z}. \tag{7.2.6}$$

However, assuming we have the non-trivial case $B_\ell \neq \varnothing$, the information provided by the system (7.2.1a) - (7.2.1d), (7.2.6) is not enough to classify the point $(\boldsymbol{y}_\ell, \boldsymbol{\sigma}_\ell, \boldsymbol{u}_\ell, \boldsymbol{p}_\ell, \boldsymbol{\mu}_\ell)$ as an S- or a C-stationary point of (4.2.4). Indeed, the system does not deliver any information on the set of biactive (active) nodes B_ℓ (A_ℓ). To overcome this difficulty, for a fixed $\epsilon > 0$, an ϵ-relaxed set of biactive nodes is introduced as

$$B_{\ell,\epsilon} := \{a_i \in Z_\ell \mid (\boldsymbol{\psi}_\ell)_i - (\boldsymbol{y}_\ell)_i \leq \epsilon\}.$$

Remark 7.2.1. The values of $\boldsymbol{y}_\ell$ in Z_ℓ are computed by means of the finite element method. Thus, on one hand, the standard definition of the biactive set $B_\ell = \{a_i \in Z_\ell \mid (\boldsymbol{\psi}_\ell)_i - (\boldsymbol{y}_\ell)_i = 0\}$ cannot be applied, because for the finite element solution $\boldsymbol{y}_\ell$ the equality $(\boldsymbol{\psi}_\ell)_i = (\boldsymbol{y}_\ell)_i$, $i \in \boldsymbol{n}_\ell^Z$ is not satisfied exactly in general. On the other hand, this implies that the choice of ϵ in the definition of the ϵ-relaxed biactive set $B_{\ell,\epsilon}$ should be related to the pointwise approximation error. The question of a proper choice of ϵ will be considered in the next chapter.

Remark 7.2.2. The concept of the ϵ-relaxed biactive set might preform poorly in case of degenerate solutions, if the degree of degeneracy is so high that in some subsets of the inactive set $\mathcal{I}$ the difference $\psi - y$ is smaller or comparable to the chosen ϵ.

Assuming that $B_{\ell,\epsilon}$ gives a good approximation of the unknown set B_ℓ (i.e., $B_\ell \approx B_{\ell,\epsilon}$), we can set $B_\ell := B_{\ell,\epsilon}$ and thus conclude that the point $(\boldsymbol{y}_\ell, \boldsymbol{\sigma}_\ell, \boldsymbol{u}_\ell, \boldsymbol{p}_\ell, \boldsymbol{\mu}_\ell)$ delivered by PDASS(LSC) satisfies the system (7.2.1a) - (7.2.1f) with

$$\boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell^B) = \mathbf{0}_{N_\ell^B}. \tag{7.2.7}$$

Independently on the sign of p_ℓ in B_ℓ, $(\boldsymbol{y}_\ell, \boldsymbol{\sigma}_\ell, \boldsymbol{u}_\ell, \boldsymbol{p}_\ell, \boldsymbol{\mu}_\ell)$ is then a C-stationary point of the discrete optimal control problem (4.2.4). If additionally

$$\boldsymbol{p}_\ell(\boldsymbol{n}_\ell^B) \geq \mathbf{0}_{N_\ell^B} \tag{7.2.8}$$

holds true, we can say that $(\boldsymbol{y}_\ell, \boldsymbol{\sigma}_\ell, \boldsymbol{u}_\ell, \boldsymbol{p}_\ell, \boldsymbol{\mu}_\ell)$ is an S-stationary point (4.2.4).

Definition 7.2.3. *(Algorithm PDASS(LSC,ϵ))*
In what follows, the algorithm PDASS(LSC) with the a posteriori detection of the biactive set via the ϵ-relaxed biactive set $B_{\ell,\epsilon}$ will be referred to as PDASS(LSC,ϵ).

Remark 7.2.4. It should be emphasized that in the post-processing step of PDASS(LSC,ϵ) the values of the discrete state function y_ℓ in $B_{\ell,\epsilon}$ are left unchanged in order to preserve the validity of the conditions in the optimality system (7.2.1). As a side effect of the convention $B_\ell := B_{\ell,\epsilon}$, one, therefore, obtains

$$\boldsymbol{\psi}_\ell(\boldsymbol{n}_\ell^A) - \boldsymbol{y}_\ell(\boldsymbol{n}_\ell^A) = \mathbf{0}_{N_\ell^A}$$

instead of

$$\boldsymbol{\psi}_\ell(\boldsymbol{n}_\ell^C) - \boldsymbol{y}_\ell(\boldsymbol{n}_\ell^C) = \mathbf{0}_{N_\ell^C}, \quad \boldsymbol{\psi}_\ell(\boldsymbol{n}_\ell^B) - \boldsymbol{y}_\ell(\boldsymbol{n}_\ell^B) \leq \epsilon \boldsymbol{I}_{N_\ell^B}.$$

It is easy to see that the complementarity constraint (7.2.1b) is nevertheless fulfilled. Thus, the conditions of C-stationarity (or S-stationarity if (7.2.8) holds) are satisfied.

Remark 7.2.5. Algorithm PDASS(LSC,ϵ) can only find S- and C-stationary points with the restriction $\boldsymbol{\mu}_\ell(\boldsymbol{n}_\ell^B) = \mathbf{0}_{N_\ell^B}$, whereas the full C- and S-stationarity systems allow $\boldsymbol{\mu}_\ell$ to be non-zero in the biactive nodes. Thus, using this algorithm, we can only find a special kind of C-stationary points, namely those that satisfy (7.2.7) for arbitrary $\boldsymbol{p}_\ell(\boldsymbol{n}_\ell^B)$, and S-stationary points satisfying (7.2.7) and (7.2.8) instead of the original condition (7.2.1g). In order to get access to the full spectrum of C- and S-stationary points of problem (4.2.4), different algorithms should be employed. For instance, the algorithms suggested in [33] and [31].

7.2.2 Remarks to the convergence of the algorithms

Chattering of the algorithm

In the numerical experiments, in case of examples with strict complementarity, the stopping criterion of the PDASS algorithm (and the PDASS(LSC,ϵ) algorithm, respectively) becomes active after a certain number of iterations. For examples with lack of strict complementarity, the algorithm sometimes enters a cycle and starts to chatter, i.e., it repeats a block of iterations none of which satisfies the stopping criterion. This effect is known in the family of primal-dual active set algorithms (see, e.g., [6]). To explain the cause of the problem, assume we have an example with $\psi_\ell \equiv 0$ that has a (S- or C-) stationary point $(\boldsymbol{y}_\ell^*, \boldsymbol{\sigma}_\ell^*, \boldsymbol{u}_\ell^*)$ with $C_\ell^* \subset A_\ell^*$ and the iterates of PDASS(LSC,ϵ) are approaching this point. At iteration k, the set $C_\ell^{(k)}$ and corresponding functions $\boldsymbol{y}_\ell^{(k)}$ and $\boldsymbol{\sigma}_\ell^{(k)}$ are very close to the exact ones. Thus, in the vicinity of the discrete biactive set $B_\ell^* = A_\ell^* \setminus C_\ell^*$ both functions $\boldsymbol{y}_\ell^{(k)}$ and $\boldsymbol{\sigma}_\ell^{(k)}$ are very small. Let $\bar{\varepsilon}$ be the accuracy of $\boldsymbol{y}_\ell^{(k)}$ and $\boldsymbol{\sigma}_\ell^{(k)}$. On iteration $k+1$, $C_\ell^{(k)} \neq C_\ell^{(k+1)}$, thus, we continue to step (iii) of the algorithm. For $a_i \in B_\ell^{(k+1)} \subseteq Z_\ell^{(k+1)}$ we set $(\boldsymbol{\sigma}_\ell^{(k+1)})_i = 0$ and from the solution of the linear system we obtain $(\boldsymbol{y}_\ell^{(k+1)})_i = \pm\,\bar{\varepsilon}$. For $a_i \in C_\ell^{(k+1)}$, we set $(\boldsymbol{y}_\ell^{(k+1)})_i = 0$ and compute $(\boldsymbol{\sigma}_\ell^{(k+1)})_i = \pm\,\bar{\varepsilon}$. The sign of the computed components of the solution will determine the zero and the strongly active set on the next iteration. However, as the components of the solution in the vicinity of the biactive set are very close to zero, their sign may be influenced, for instance, by round-off errors. As a result, the algorithm might start to chatter.

Note that for certain examples with lack of strict complementarity and $p|_{\mathcal{B}} = 0$, chattering did not take place and the algorithm was invariably terminated by the regular stopping criterion. For those examples where chattering occurred, an additional stopping criterion suggested in [41] was used: from the block of repeated iterations, the one feasible in the sense of

$$\boldsymbol{\sigma}_\ell^{(k)} \geq \mathbf{0}_{N_\ell} \tag{7.2.9}$$

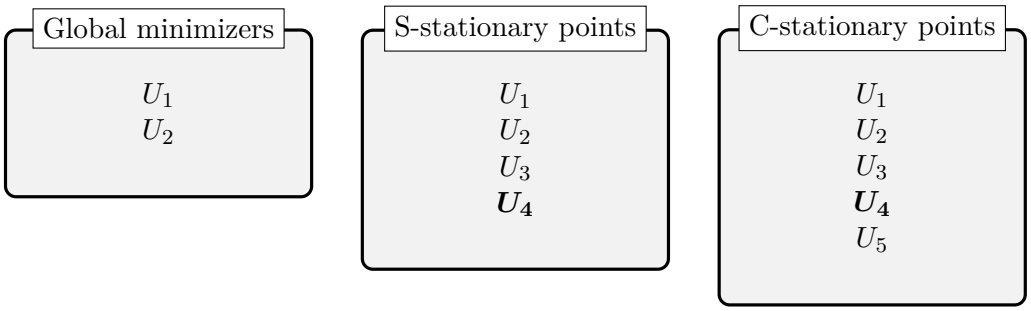
is picked. For the example considered, at most of the uniform (adaptive) mesh levels there was only one element of the cycle that satisfied this condition. Still, at some mesh levels, none of the iterations from the repeated block satisfied (7.2.9). In such cases, the solution algorithm was terminated after reaching the number of maximum admissible iterations k_{max}. Note that when the algorithm is terminated by the additional stopping criterion (7.2.9) or after reaching $k = k_{max}$, the computed solution does not satisfy the system of stationarity conditions exactly.

Stationary points and global solutions

In general there is no guarantee that the numerical algorithm used will converge to the same stationary point on every mesh. In practice, however, for the same set of data convergence to one and the same solution was observed in all simulations.

In the following schematic example, we look at the different types of discrete stationary points we might find by means of our numerical algorithms and, once again, emphasize their relation to each other as well as to the global minimizers. Furthermore, the place of the solutions with strict complementarity in the hierarchy of stationary points is addressed.

Example 7.2.6. Using the notation of Section 5.3, let SOL_ℓ^S denote the set of S-stationary points of the discrete optimal control problem. As depicted in the scheme below, suppose we have an example with $\mathrm{SOL}_\ell^{glob} = \{U_1, U_2\}$, $\mathrm{SOL}_\ell^S = \{U_1, U_2, U_3, \boldsymbol{U_4}\}$, and $\mathrm{SOL}_\ell^C = \{U_1, U_2, U_3, \boldsymbol{U_4}, U_5\}$. The bold character marks the solution satisfying the strict complementarity condition.



The hierarchy "global minimizer ⇒ S-stationary point ⇒ C-stationary point" implies:

- the global minimizers U_1, U_2 must enter both, the set of S-stationary points and the set of C-stationary points;
- the S-stationary points U_1, U_2, U_3, and $\boldsymbol{U_4}$ are found in SOL_ℓ^C.

Using an algorithm that looks for C-stationary or S-stationary points of the discrete problem, we might converge to $U_5 \in \mathrm{SOL}_\ell^C \setminus \mathrm{SOL}_\ell^{glob}$ or $U_3 \in \mathrm{SOL}_\ell^S \setminus \mathrm{SOL}_\ell^{glob}$.

Stationary point $\boldsymbol{U_4}$ satisfies the strict complementarity condition. Strict complementarity eliminates the differences between C- and S-stationarity concepts, thus, any stationary point with this property is a C- and an S-stationary point, but not necessary a global minimizer. In our example, U_5 *must* have lack of strict complementarity, whereas the stationary points U_1 - U_3 could, in principle, have either strict complementarity or lack of strict complementarity. Applying an algorithm that looks exclusively for stationary points with strict complementarity to this example, we a priori exclude the global minimizers from our search.

7.3 Error indication, marking, refinement

In step ESTIMATE of the adaptive algorithm 7.1, the components of the residual-type a posteriori error estimate (6.1.2) are computed for the discrete stationary point provided by algorithm PDASS (PDASS(LSC,ϵ)). Estimate (6.1.2) not only gives global upper and lower bounds to the discretization error, but also tells us how the error is distributed over the computational domain. This local information allows to construct local error indicators that lie at the heart of any adaptive algorithm. Based on local error indicators, the adaptive algorithm decides where to place new degrees of freedom in order to adjust the mesh to the structure of a particular problem.

On every mesh $\mathcal{T}_\ell$, an elementwise error indicator $ind_\ell \in W_\ell$ is computed according to

$$ind_\ell := ind(\eta_\ell) + ind(osc_\ell), \tag{7.3.1}$$

where $ind(\eta_\ell) \in W_\ell$ is the indicator composed of the values of the element residuals and weighted sums of the edge residuals on edges adjacent to the elements, i.e.,

$$ind(\eta_\ell) := \eta_\ell^{y,T} + \eta_\ell^{y,E} + \eta_\ell^{p,T} + \eta_\ell^{p,E},$$

with $\eta_\ell^{y,T}$, $\eta_\ell^{y,E}$, $\eta_\ell^{p,T}$, $\eta_\ell^{p,E} \in W_\ell$, such that for any $T \in \mathcal{T}_\ell$,

$$\eta_\ell^{y,T} := \sum_{T \subseteq \mathcal{Z}_\ell} \eta_T(y)\chi(T), \qquad \eta_\ell^{y,E} := \sum_{T \subseteq \mathcal{Z}_\ell} \left(\frac{1}{2} \sum_{E \in \mathcal{E}_\ell(T) \cap \mathcal{E}_{\mathcal{Z}_\ell}} \eta_E(y) \right) \chi(T),$$

$$\eta_\ell^{p,T} := \sum_{T \subseteq \mathcal{I}_\ell} \eta_T(p)\chi(T), \qquad \eta_\ell^{p,E} := \sum_{T \subseteq \mathcal{I}_\ell} \left(\frac{1}{2} \sum_{E \in \mathcal{E}_\ell(T) \cap \mathcal{E}_{\mathcal{I}_\ell}} \eta_E(p) \right) \chi(T).$$

The indicator $ind(\mathrm{osc}_\ell) \in W_\ell$ is based on the combined data oscillation from the reliability and the efficiency estimates,

$$ind(osc_\ell) := \sum_{T \in \mathcal{T}_\ell} osc_T(u^d)\chi(T) + \sum_{T \subseteq \mathcal{Z}_\ell} osc_T(f)\chi(T) + \sum_{T \subseteq \mathcal{I}_\ell} osc_T(y^d)\chi(T).$$

In step MARK, the elements of $\mathcal{T}_\ell$ are marked for refinement, using the elementwise error indicator (7.3.1) and the so-called bulk criterion (see, e.g., [20, 50]): for a fixed parameter $\Theta \in (0,1)$, a set of elements $\mathcal{M}_{\ell,\Theta} \subset \mathcal{T}_\ell$ is selected such that

$$\Theta \sum_{T \in \mathcal{T}_\ell} ind_\ell|_T \leq \sum_{T \in \mathcal{M}_{\ell,\Theta}} ind_\ell|_T \,. \tag{7.3.2}$$

The bulk criterion is realized by a greedy algorithm (cf., e.g., [35]).

In step REFINE, we use the standard Matlab PDE Toolbox routine *refinemesh*. This routine offers two mesh refinement algorithms: "longest" and "regular". According to Matlab documentation, both algorithms follow the steps specified below:

Initialization:	(i)	receive the set of triangles to be refined from the input;
	(ii)	divide all the edges of the selected triangles in half ("regular"),
		or divide the longest edge in half ("longest");
Iteration:	(iii)	divide the longest edge of any triangle that has a divided edge;
	(iv)	stop if no further edges have to be devided,
		otherwise go to (iii).

Remark 7.3.1. In the process of investigating the properties of the adaptive algorithm, more types of error indicators have been considered, for instance:

- (a) indicators including the information delivered by the heuristic consistency error estimates;
- (b) indicators that include the discrete free boundaries $\mathcal{F}_\ell(y_\ell)$, $\mathcal{F}_\ell(\sigma_\ell)$ into the refinement process on each iteration of the adaptive algorithm;
- (c) a combination of (a) and (b).

However, as (a) did not have pronounced influence on the performance of the adaptive algorithm, whereas (b) only worsened the results, the documentation of these experiments is omitted in this work.

7.4 Comparative algorithm: uniform mesh refinement

In the numerical tests, the outcome of the adaptive algorithm 7.1 will be compared to that of the following "uniform" algorithm.

Initialization:		set $\ell = 0$, choose $\mathcal{T}_\ell$, SOLVE $\rightarrow$ ESTIMATE on $\mathcal{T}_\ell$,
		check stopping criterion;
Iteration:	(i)	set $\ell = \ell + 1$, MARKALL $\rightarrow$ REFINE ;
	(ii)	SOLVE $\rightarrow$ ESTIMATE on $\mathcal{T}_\ell$;
	(iii)	check stopping criterion.

Algorithm 7.5: The uniform algorithm.

In step MARKALL all elements of the current triangulation $\mathcal{T}_\ell$ are marked for refinement. Otherwise, Algorithm 7.5 is identical to Algorithm 7.1 and uses the same initial mesh and stopping criterion.

8 Numerical experiments

In this chapter, the results of the numerical experiments for problems with and without strict complementarity are presented. They illustrate the performance of the suggested finite element approximation. All the experiments were carried out in MATLAB 7.9.0 (R2009b), with machine precision $\epsilon_M = 2.2204e\text{-}16$. For the solution of the systems of linear equations in step (iii) of the algorithms PDASS and PDASS(LSC,ϵ), the MATLAB direct solver for linear systems with sparse matrices was employed.

8.1 Convergence rates

Let us first introduce a formal framework for the evaluation of the speed of convergence of the finite element approximations computed on sequences $\{\mathcal{T}_\ell\}_{\mathbb{N}}$ of (uniform or adaptive) simplicial meshes with associated degrees of freedom $\mathrm{DOF}_\ell := \mathcal{N}_\ell$.

In the theory of finite element methods, the asymptotic behavior of the approximation error err_ℓ associated with finite element solutions computed on uniformly refined meshes is predicted by the so-called a priori error estimates. These estimates usually claim that

$$err_\ell \leq C h_\ell^\alpha, \tag{8.1.1}$$

where h_ℓ is the maximal diameter of the finite element mesh (cf. Paragraph 6.6.1), C is a nonnegative constant independent on the mesh size h_ℓ, and $\alpha > 0$. The estimate (8.1.1) suggests to measure the decay of the error err_ℓ in terms of the mesh size h_ℓ.

Definition 8.1.1. *(Convergence rate in terms of h_ℓ)*
If there exists a real number $\alpha > 0$ such that

$$err_\ell = O(h_\ell^\alpha), \tag{8.1.2}$$

it is referred to as the convergence rate of err_ℓ with respect to the mesh size h_ℓ.

When piecewise linear finite element approximations are used for the discretization of a problem with sufficiently smooth data (so that the solution of the problem possesses sufficient regularity) and if err_ℓ represents the $H^1(\Omega)$-norm of the approximation error, one typically obtains convergence rates $\alpha = 1$, whereas for problems with less regular solutions usually only $\alpha < 1$ can be achieved.

Remark 8.1.2. Let $y \in V$ stand for the solution of an elliptic variational inequality of obstacle type satisfying the assumptions of Lemma 2.2.6. It is analytically verified (see, e.g., [18], Chapter 5) that the approximation error $err_\ell = \|y - y_\ell\|_{1,\Omega}$ associated with

a finite element approximation $y_\ell \in V_\ell$ exhibits an asymptotic convergence rate $\alpha = 1$. For finite element discretization of the optimally controlled elliptic obstacle problem, no such result is available in the literature. However, in the numerical experiments convergence rates $\alpha \approx 1$ were obtained for examples with sufficiently smooth data.

For the adaptively refined meshes, we usually have $h_\ell \geq const. > 0$ for $\ell \to \infty$. Thus, convergence expectations in terms of h_ℓ are meaningless. In this case, it is appropriate to measure the decay of the error err_ℓ in terms of the degrees of freedom DOF_ℓ provided by the finite element mesh $\mathcal{T}_\ell$ (see, e.g., [50, 16]). Observing that for the uniformly refined meshes there holds $h_\ell = O(\mathrm{DOF}_\ell^{-1/2})$, the asymptotic identity (8.1.2) obviously admits the following reformulation in terms of DOF_ℓ:

$$err_\ell = O(\mathrm{DOF}_\ell^{-\beta}), \quad \beta := \alpha/2 > 0. \tag{8.1.3}$$

Definition 8.1.3. *(Convergence rate in terms of* DOF_ℓ*)*
If there exists a real number $\beta > 0$ *such that*

$$err_\ell = O(\mathrm{DOF}_\ell^{-\beta}), \tag{8.1.4}$$

it is referred to as the convergence rate of err_ℓ *with respect to the number of degrees of freedom* DOF_ℓ*.*

Remark 8.1.4. On a double logarithmic scale $x := \log_{10}(\mathrm{DOF}_\ell)$, $y := \log_{10}(err_\ell)$, the dependence $err_\ell = C\,\mathrm{DOF}_\ell^{-\beta}$, $C > 0$ is represented by the affine function

$$y = -\beta\, x + \tilde{C}, \quad \tilde{C} = \log_{10}(C).$$

Remark 8.1.5. For the sequences of uniformly refined meshes, the convergence in terms of DOF_ℓ (h_ℓ) can be found from the relation $\beta = \alpha/2$ as soon as the convergence in terms of h_ℓ (DOF_ℓ) is determined. In general, α cannot be determined in case of the adaptively refined meshes.

For problems with regular solutions, the approximations computed on the sequences of adaptively refined meshes usually show the same asymptotic behavior as the approximations obtained on the sequences of uniform meshes. Differences in the asymptotic behavior are observed in application to problems with less regular solutions. The adaptive mesh refinement aims to distribute the degrees of freedom DOF_ℓ of the mesh $\mathcal{T}_\ell$ in the computational domain Ω in such a way that the approximation error err_ℓ is minimized. For problems with less regular solutions, this approach allows to achieve asymptotical convergence rates $\beta = 1/2$ for the approximation error in the $H^1(\Omega)$-norm, whereas the uniform refinement can only deliver $\beta < 1/2$.

In order to analyze the asymptotical behavior of the finite element approximations in practice, the notion of *experimental convergence rates* in terms of DOF_ℓ is introduced.

Definition 8.1.6. *(Experimental convergence rates in terms of DOF_ℓ)*
For $\ell \in \mathbb{N}$, we refer to

$$\beta_\ell(err_\ell) := \frac{\log_{10}(err_{\ell-1}/err_\ell)}{\log_{10}(\mathrm{DOF}_\ell \,/\, \mathrm{DOF}_{\ell-1})}$$

as the experimental convergence rate of err_ℓ in terms of DOF_ℓ.

In the definitions of the (experimental) convergence rates, the notion of the approximation error err_ℓ can be substituted by any other quantity of interest associated with the finite element solution whose asymptotic behavior with the growth of DOF_ℓ we want to observe. For example, one can consider the convergence rates of the error estimator η_ℓ and its components. Hence, from now on we allow err_ℓ to represent any quantities associated with the finite element solution.

In the subsequent numerical examples, the behavior of err_ℓ will be visualized using linear interpolation of the values $err_\ell(\mathrm{DOF}_\ell)$ on a double logarithmic scale. In this case, the numbers $\beta_\ell(err_\ell)$ correspond to the absolute value of the negative slopes of the lines connecting $\log_{10}(err_{\ell-1})$ and $\log_{10}(err_\ell)$, i.e., they shows how fast we descend from $\log_{10}(err_{\ell-1})$ to the new value $\log_{10}(err_\ell)$. These lines will be compared both for uniform refinement and adaptive refinement. In case of regular solutions, we should expect the slopes associated with the approximation errors (the error estimator) to be approximately the same, whereas for less regular solutions the slope for adaptive refinement is expected to be steeper than in case of uniform refinement.

8.2 Choice of ϵ in PDASS(LSC, ϵ)

As pointed out in Remark 7.2.1, the value of ϵ in the definition of the $\epsilon-$relaxed biactive set $B_{\ell,\epsilon}$ used by the algorithm PDASS(LSC, ϵ) is related to the pointwise approximation error $e_{\ell,y} = y - y_\ell$. Assuming that $e_{\ell,y}$ is uniformly distributed over the computational domain and $\|e_{\ell,y}\|_{0,\Omega}$ decays with the asymptotic convergence rate $\beta = 1/2$, set

$$\epsilon := \gamma^{-1} DOFs(\ell)^{-1/2},$$

where $\gamma > 0$ is the correction parameter. In the numerical experiments, the value $\gamma = 50$ was used.

The same parameter ϵ is used to identify the type of stationary points delivered by the algorithm PDASS(LSC, ϵ): a point is classified as a S-stationary point if

$$p(\boldsymbol{n}_\ell^B) \geq -\epsilon \boldsymbol{I}_{N_\ell^B}.$$

8.3 Examples

In this section, a selection of numerical test problems with and without strict complementarity is presented. These test problems will illustrate the behavior of the finite element approximations in practice. Four of the six problems have a known analytical solution, which allows to compare the approximated (estimated) quantities with their exact counterparts.

8.3.1 Examples with strict complementarity

Example 1

Let us consider $A = -\Delta$ on the L-shaped domain $\Omega = (-2,2)^2 \setminus ([0,2] \times [-2,0])$.

In polar coordinates (r, φ), let an auxiliary function $\gamma \in C^1(\overline{\Omega})$ be defined according to

$$\gamma(r) := \begin{cases} 0 & , \ r \geq \bar{r} := 2 \\ 0.25r^3 - 0.75r^2 + 1, & \text{else} \end{cases} .$$

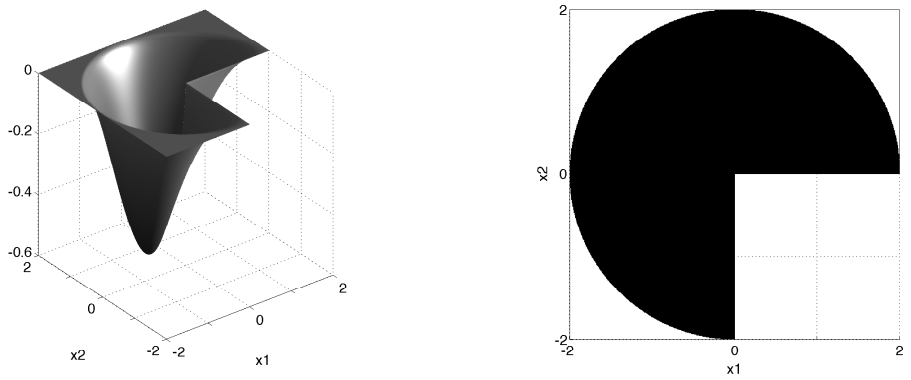


Figure 8.1: Example 1. Optimal state y^* (left) and the inactive set $\mathcal{I}^*$, marked in black (right).

Given

$$y^*(r,\varphi) = -\gamma(r) r^{2/3} \sin\left(\frac{2}{3}\varphi\right), \quad \sigma^*(r) = \begin{cases} 1, & r \geq \bar{r} \\ 0, & \text{else} \end{cases} ,$$

it can be verified analytically that the triple $(y, \sigma, u) = (y^*, \sigma^*, y^*) \in V \times L^2(\Omega) \times L^2(\Omega)$ with the adjoint state $p = y^*$ and the multiplier $\mu = \sigma^*$ is a stationary point with strict complementarity of (3.1.1) with respect to the data

$$\begin{aligned} &y^d = \mu - \Delta p + y, \quad u^d \equiv 0, \quad \alpha = 1, \\ &f = \sigma - \Delta y - p, \quad \psi \equiv 0. \end{aligned}$$

Further, there holds $\mathcal{I}^* = \{(r, \varphi) \,|\, r \in (0, \bar{r}),\, \varphi \in (0, 3\pi/2)\}$, $\mathcal{Z}^* = \mathcal{I}^*$. The state y^* and the inactive set $\mathcal{I}^*$ are displayed in Figure 8.1.

Example 2

A variation of Example 1 with $\bar{r} = 1/2$ and

$$\gamma(r) := \begin{cases} 0 & , \; r \geq \bar{r} \\ 16r^3 - 12r^2 + 1, & \text{else} \end{cases} .$$

The state function y^* and the inactive set $\mathcal{I}^*$ are displayed in Figure 8.2.

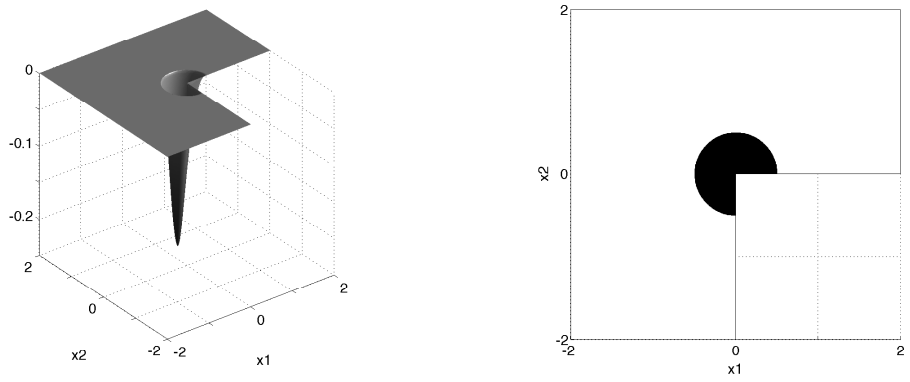


Figure 8.2: Example 2. Optimal state y^* (left) and inactive set $\mathcal{I}^*$, marked in black (right).

Example 3

For this example let us again consider the operator $A = -\Delta$, this time on the domain $\Omega = (0, 1)^2 \setminus ([0.75, 1]^2)$. The example is defined by the data

$$\begin{aligned} &y^d(x_1, x_2) = -|x_1 x_2 - 0.5| + 0.25, \quad u^d \equiv 0, \quad \alpha = 0.01, \\ &f = y^d, \quad \psi \equiv 0 \end{aligned}$$

(cf. [45], p. 66). For this example no analytical solution is available. The numerical approximation of the optimal solution is shown in Figure 8.3. The optimal state function exhibits a very flat transition into the active set (degenerate solution) which makes this example challenging for the PDASS algorithm.

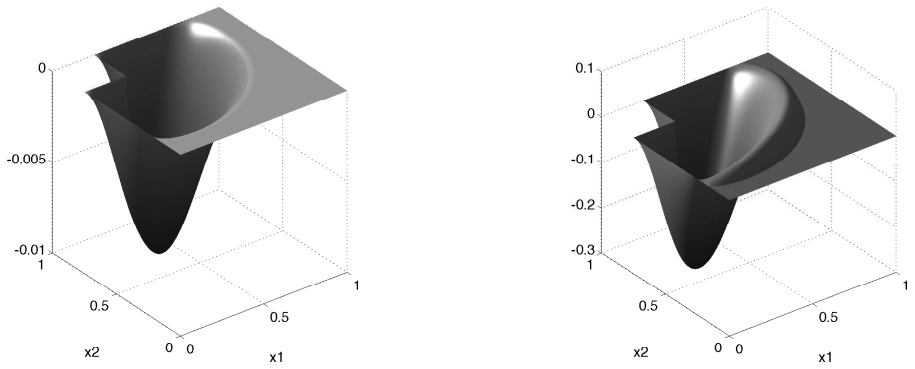


Figure 8.3: Example 3. Discrete state y_ℓ (left) and discrete control u_ℓ (right), mesh level $\ell = 10$ of uniform algorithm.

8.3.2 Examples with lack of strict complementarity

Example 4

This example was originally constructed in [33]. Consider a unit square geometry $\Omega = (0,1)^2$. With the help of auxiliary functions

$$z_1(x_1) = -4096\,x_1^6 + 6144\,x_1^5 - 3072\,x_1^4 + 512\,x_1^3,$$
$$z_2(x_2) = -244.140625\,x_2^6 + 585.9375\,x_2^5 - 468.75\,x_2^4 + 125\,x_2^3$$

we define

$$y^*(x_1,x_2) = \begin{cases} -z_1(x_1)z_2(x_2), & (x_1,x_2) \in (0,0.5)\times(0,0.8) \\ 0 \qquad\qquad\quad , & \text{else} \end{cases},$$
$$\sigma^*(x_1,x_2) = 2\max(0, -|x_1 - 0.8| - |(x_2 - 0.2)x_1 - 0.3| + 0.35).$$

The triple $(y,\sigma,u) = (y^*,\sigma^*,y^*) \in V \times L^2(\Omega) \times L^2(\Omega)$ with the multipliers $(p,\mu) = (y^*,\sigma^*)$ is an S-stationary point for problem (3.1.1) with the data

$$\alpha = 1, \quad \psi \equiv 0, \quad u^d \equiv 0,$$
$$f = \sigma - \Delta y - p,$$
$$y^d = \mu - \Delta p + y.$$

It is easy to see that in this example

$$\mathcal{I} = \{(x_1,x_2)\,|\,(x_1,x_2) \in (0,0.5)\times(0,0.8)\},$$
$$\mathcal{C} = \{(x_1,x_2)\,|\,|x_1 - 0.8| + |(x_2 - 0.2)x_1 - 0.3| \le 0.35\},$$

and, consequently, $\mathcal{B} = \Omega \setminus (\mathcal{I} \cup \mathcal{C}) \neq \varnothing$. The exact state y^* and Lagrangian multiplier σ^* are shown in Figure 8.4. The exact inactive set $\mathcal{I}^*$ and strongly active set $\mathcal{C}^*$ in Figure 8.5.

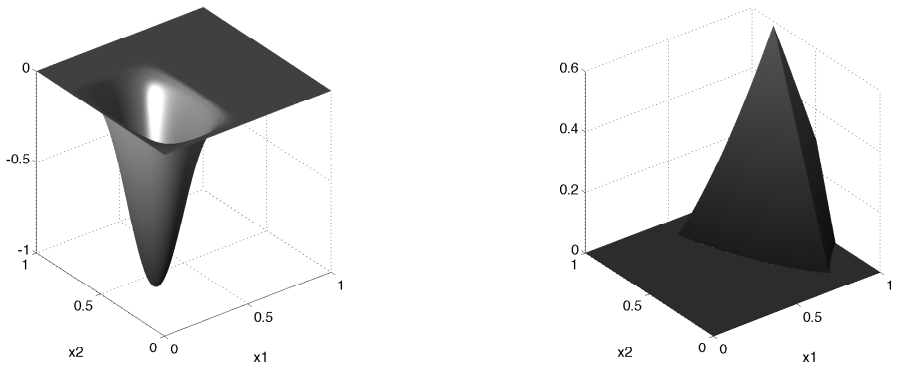


Figure 8.4: Example 4. Optimal state y^* (left) and corresponding Lagrangian multiplier σ^* (right).

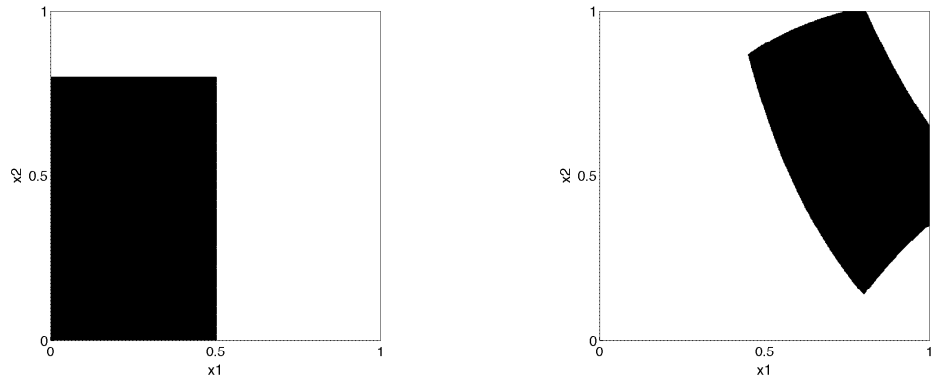


Figure 8.5: Example 4. In black: the exact inactive set $\mathcal{I}^*$ (left) and the exact strongly active set $\mathcal{C}^*$ (right).

Example 5

This example is very similar to Example 2. For $A = -\Delta$ and the L-shaped domain $\Omega = (-2,2)^2 \setminus ([0,2] \times [-2,0])$, the optimal state function y^* is chosen identical to that of Example 2, whereas the multiplier σ^* is defined according to

$$\sigma^*(r) = \begin{cases} 1, & r \geq \bar{r} = 2 \\ 0, & \text{else} \end{cases} .$$

The triple $(y, \sigma, u) = (y^*, \sigma^*, y^*)$ with $(p, \mu) = (y^*, \sigma^*)$ is an S-stationary point for problem (3.1.1) with the data

$$\begin{aligned} &\alpha = 1, \quad \psi \equiv 0, \quad u^d \equiv 0, \\ &f = \sigma - \Delta y - p, \\ &y^d = \mu - \Delta p + y. \end{aligned}$$

The inactive set $\mathcal{I}^*$ and the strongly active set $\mathcal{C}^*$ are displayed in Figure 8.6.

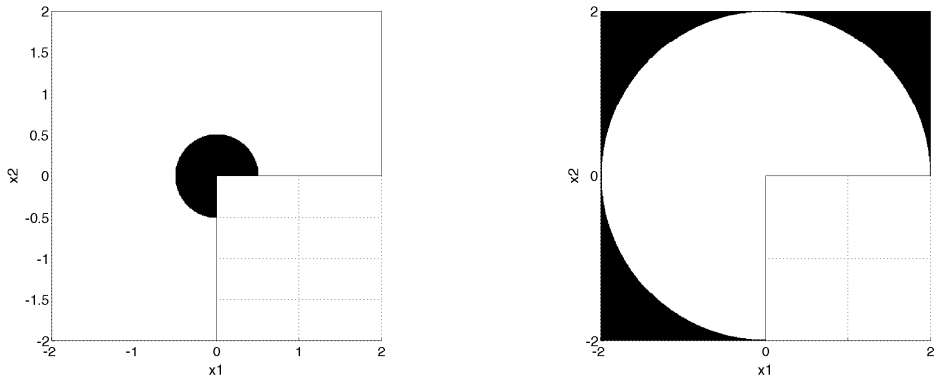


Figure 8.6: Example 5. In black: the exact inactive set $\mathcal{I}^*$ (left) and the exact strongly active set $\mathcal{C}^*$ (right).

Example 6

Let us consider $A = -\Delta$ on the domain displayed in Figure 8.8. This example is defined by the data

$$\begin{aligned} &y^d(x_1, x_2) = 5x_1 + x_2 - 1, \quad u^d \equiv 0, \quad \alpha = 1, \\ &f(x_1) = x_1 - 0.5, \quad \psi \equiv 0 \end{aligned}$$

(cf. [41], p. 363). For this example no analytical solution is available. The numerical approximation of the optimal solution at mesh level $\ell = 9$ of the uniform algorithm is shown in Figures 8.7 and 8.8.

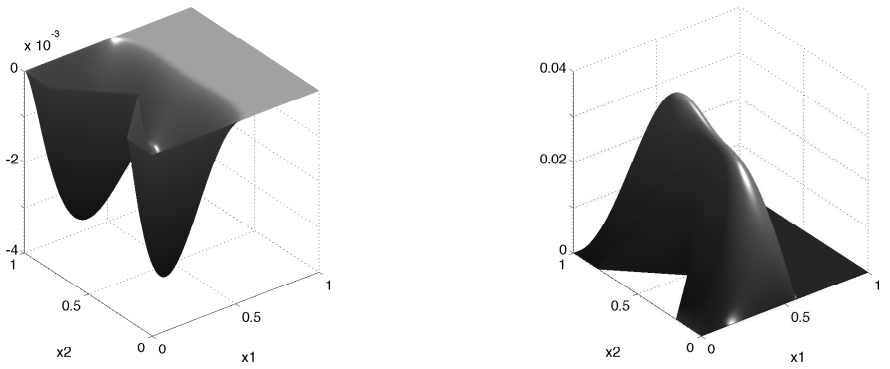


Figure 8.7: Example 6. Discrete state y_ℓ (left) and discrete control u_ℓ (right), mesh level $\ell = 9$ of uniform algorithm.

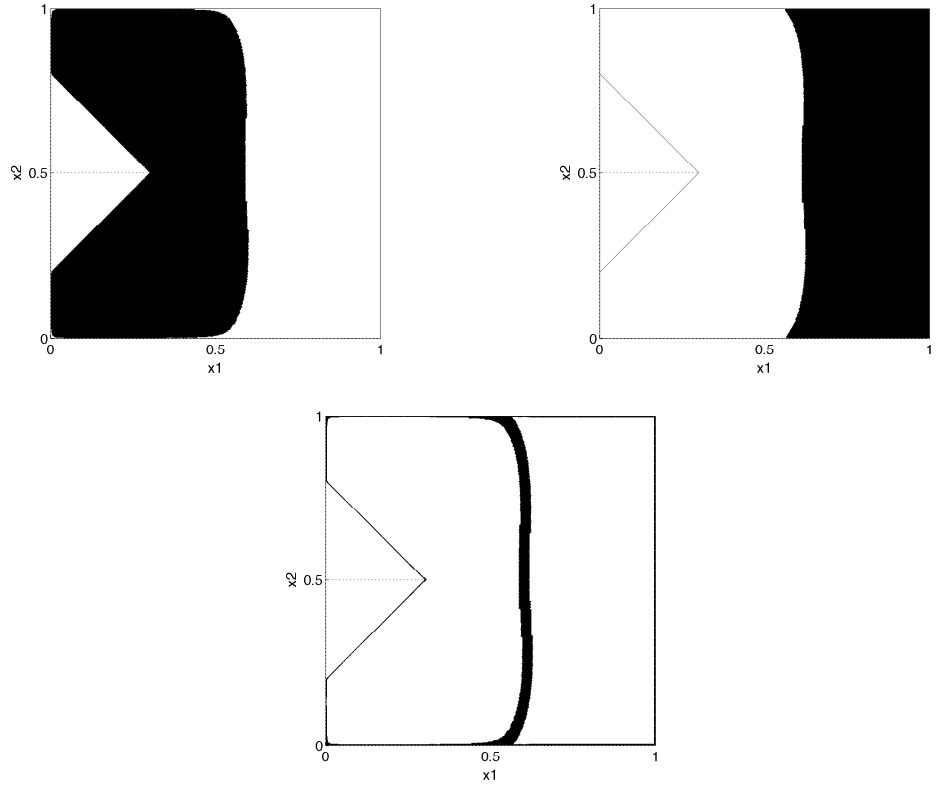


Figure 8.8: Example 6. In black: discrete inactive $\mathcal{I}_\ell$ (top left), strongly active $\mathcal{C}_\ell$ (top right), and biactive set $\mathcal{B}_\ell$ (bottom) for $\ell = 9$, mesh level $\ell = 9$ of uniform algorithm.

8.4 Numerical results

In this section, a detailed documentation of the numerical simulations is given that were performed in order to test the adaptive algorithm 7.1 in comparison to the uniform algorithm 7.5. Furthermore, the behavior of the heuristic estimates of the consistency error associated with the efficiency and the reliability estimate as well as the quality of the approximation of the discrete sets is analyzed for both, the uniform and the adaptive mesh refinement.

The sequences of adaptive finite element approximations are computed using the value $\Theta = 0.55$ in step MARK and the algorithm "regular" (cf. Section 7.3) in step REFINE of Algorithm 7.1. The finite element approximations on the uniform meshes are computed employing the algorithm "longest" in step REFINE of Algorithm 7.5.

The test problems introduced in the previous section can be divided into two groups:

- problems that have a stationary point with strict complementarity (Examples 1, 2, and 3);
- problems that have an S-stationary point with lack of strict complementarity that satisfies $\mu|_{\mathcal{B}} = 0$ (Examples 4, 5, and 6).

In step SOLVE of both, the adaptive and the uniform loops, the PDASS algorithm is used to solve the problems with strict complementarity, whereas PDASS(LSC,ϵ) is employed for problems with lack of strict complementarity. Note that in case of Examples 3 and 6 chattering of the solver is observed. In case of Example 3, the cycling is due to the degeneracy of the solution. When none of the iterations in the cycle satisfies the additional stopping criterion (7.2.9), the solver is terminated after reaching the maximum admissible number of iterations $k_{max} = 202$. The effect of cycling on the performance of the error estimate will be addressed in the next paragraph.

8.4.1 Adaptive meshes

Figures 8.9 and 8.10 show adaptive meshes on selected levels ℓ for Example 1, 2, 3, 5, and 6 with a zoom into the vicinity of the singularity of the solution. As expected, adaptive mesh refinement leads to a high resolution of the finite element mesh in the inactive set and, in particular, around the singularity. Figure 8.10 (top) corresponds to Example 4 whose solution is smooth, it shows the adaptively generated mesh at levels $\ell = 10$ and $\ell = 7$. Note that Figure 8.10 shows that the strongly active set $\mathcal{C}^*$ is not well resolved by the adaptive mesh refinement (cf. Figures 8.5, 8.6, and 8.8 illustrating strongly active set of Examples 4, 5, and 6, respectively).

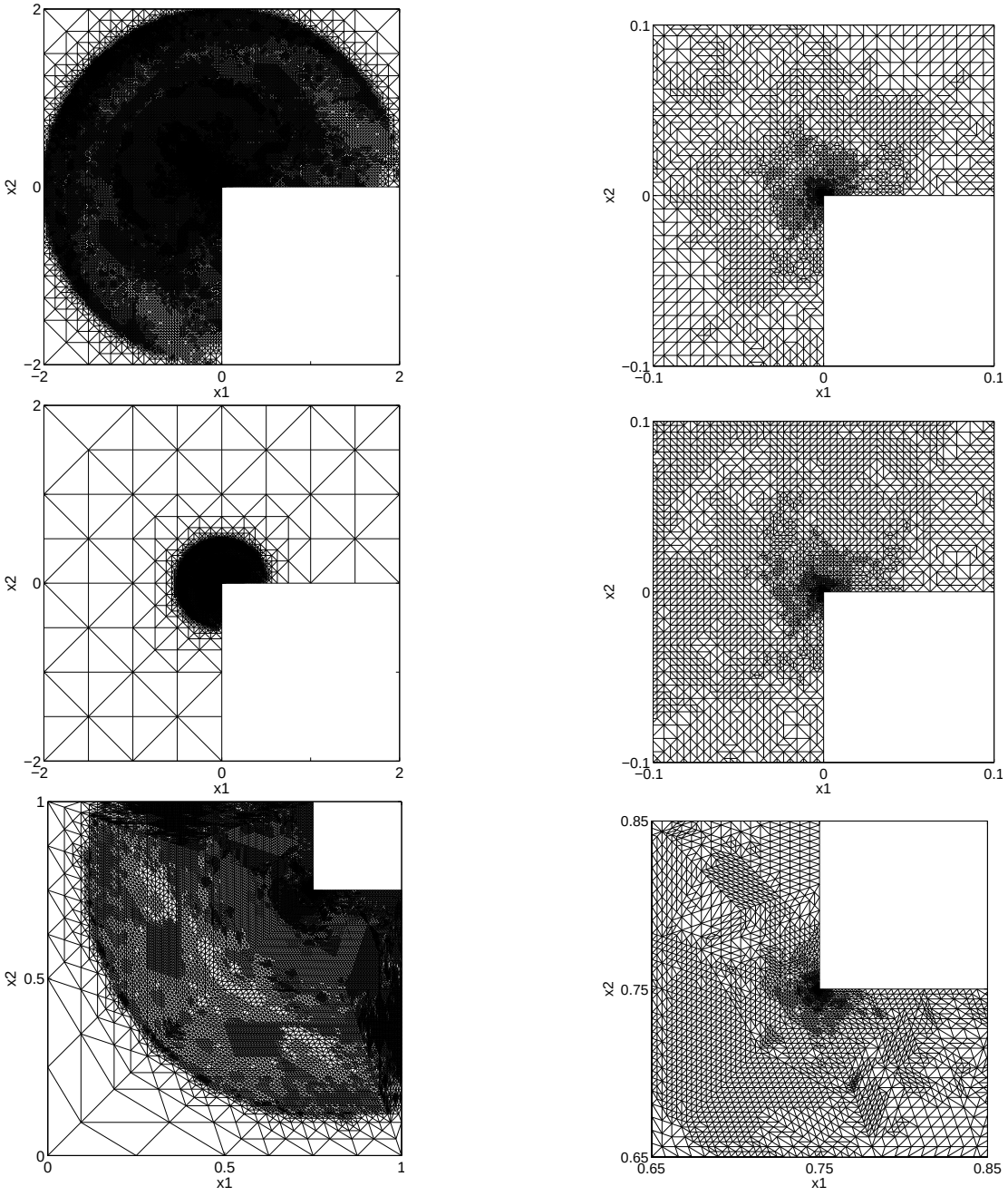


Figure 8.9: Example 1 (top), Example 2 (middle), Example 3 (bottom). Meshes $\ell = 10$, $\ell = 11$, $\ell = 8$, respectively, (left) and zoom into the vicinity of the singularity (right).

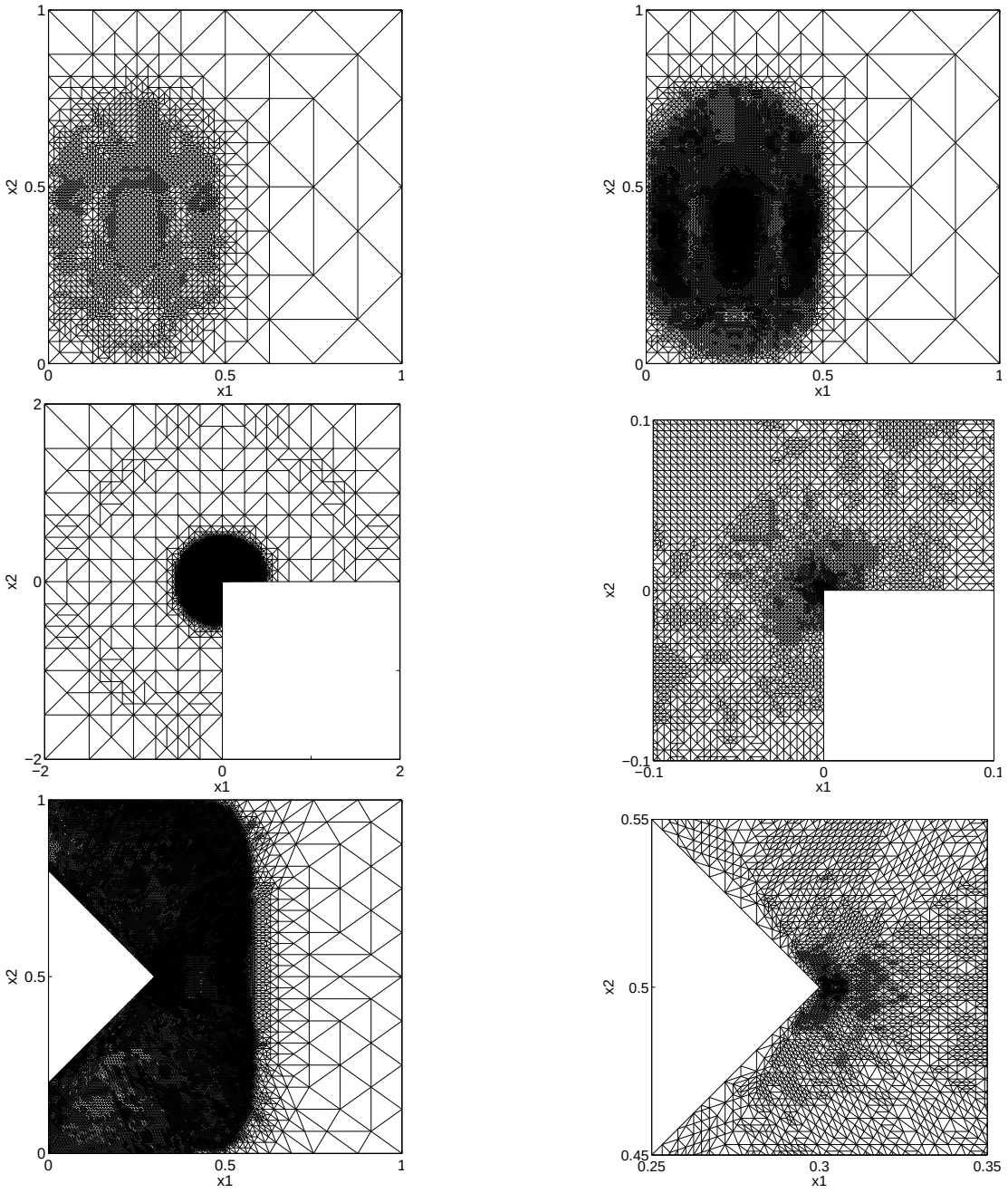


Figure 8.10: Example 4 (top): meshes $\ell = 7$ (left) and $\ell = 10$ (right). Example 5 (middle), Example 6 (bottom): meshes $\ell = 11$, $\ell = 10$, respectively, (left) and zoom into the vicinity of the singularity (right).

8.4.2 Convergence history and experimental convergence rates

The convergence of the finite element approximations is analyzed in terms of the error estimator η_ℓ. When the analytical solution of the test problem is available, the behavior of both, the total error $|||e_\ell|||$ and error estimator η_ℓ, is examined.

Remark 8.4.1. The values of the separate components of the total error and the error estimator as well as the associated experimental convergence rates for uniform and adaptive refinement can be found in the collection of tables presented in Appendix A.

Figures 8.11 and 8.12 illustrate the decrease of the total error $|||e_\ell|||$ and/or of the estimator η_ℓ as a function of the DOF on a double logarithmic scale, as well as the experimental convergence rates associated with the error estimator $\beta_\ell(\eta_\ell)$ (when available, also $\beta_\ell(|||e_\ell|||)$ as a function of the DOF on a semi-logarithmic scale.

The plots corresponding to the examples with known exact solution show that the error estimator η_ℓ indeed provides an upper bound to the total error $|||e_\ell|||$ with essentially the same convergence rate.

The benefit of the adaptive algorithm can be seen from the behavior of the experimental convergence rates for the examples with less regular solutions (these are all the examples except for Example 4). For Examples 1, 2, and 5, the experimental convergence rates corresponding to the adaptive solutions asymptotically approach the value of 0.5, whereas the uniform refinement leads to convergence rates $\beta_\ell(\eta_\ell) \approx \beta_\ell(|||e_\ell|||) \approx 0.4$.

The solution of Example 4 is smooth, thus, the uniform refinement is already optimal. As it can be seen in Figure 8.12 top right, similar experimental convergence rates are obtained for uniform and adaptive mesh refinement. Though in case of uniform refinement we observe a more pronounced oscillation around the value 0.5.

Chattering of the solver

In case of Examples 3 and 6, the experimental convergence rates are oscillating in the vicinity of the value 0.5 for adaptive refinement and of 0.4 for uniform refinement. This unstable behavior of the experimental convergence rates is caused by chattering of the solver. However, also in case of these two examples one can see the advantage of the adaptive mesh refinement comparing both, the plots of the experimental convergence rates and the plots illustrating the decay of the error estimator.

The documentation of the performance of PDASS (PDASS(LSC,ϵ)) within the iterations of the uniform and adaptive algorithms for Examples 3 and 6 can be found in Table A.3.5 and Tables A.6.5, A.6.6, respectively. In the tables, the levels ℓ of the uniform (adaptive) algorithm where chattering occurs can be identified by the number of iterations k'_ℓ of PDASS (PDASS(LSC,ϵ)).

- $k'_\ell \leq 151$ - *no chattering*, termination by the regular stopping criterion.
- $152 \leq k'_\ell \leq 201$ - *chattering*, termination by the additional stopping criterion (7.2.9).

- $k'_\ell = 202$ - *chattering*, from the repeated block of iterations of PDASS (PDASS(LSC,ϵ)) none satisfies the additional stopping criterion (7.2.9). The solver is terminated after reaching the maximum admissible number of iterations $k_{max} = 202$.

As it can be seen from the tables, at some mesh levels chattering did not occur. In particular, in case of uniform mesh refinement, we observe no chattering

- at level $\ell = 3$ for Example 3,
- at level $\ell = 0$ for Example 6.

For the adaptive mesh refinement, there is no chattering

- at levels $\ell \in \{1, 3, 4\}$ for Example 3,
- at levels $\ell \in \{0, 1, 8, 11, 12\}$ for Example 6.

Thus, it seems that the adaptive mesh refinement to some extent helps to avoid chattering of the solver.

Remark 8.4.2. When chattering occurs, the algorithm is terminated either by the additional stopping criterion (7.2.9), or, in the worst case, by reaching the number of maximum admissible iterations k_{max}. Both scenarios lead to a situation where the discrete solution provided by the algorithm violates some of the optimality conditions (cf. Paragraph 7.2.2). This inaccuracy, though not visible in the behavior of the error estimator (see Figures 8.11 and 8.12), can significantly influence the behavior of other quantities reflecting finer properties of the approximations such as

- the experimental convergence rates,
- the consistency errors $e^c_{\ell,rel}$, $e^c_{\ell,eff}$ and their heuristic estimates,
- the quantities $e^{evd}_{\ell,\mathcal{A}}$, $e^{dva}_{\ell,\mathcal{A}}$, $e^{eva}_{\ell,\mathcal{A}}$ and $e^{evd}_{\ell,\mathcal{C}}$, $e^{dva}_{\ell,\mathcal{C}}$, $e^{eva}_{\ell,\mathcal{C}}$ reflecting the quality of the approximation of the discrete sets and related heuristic set approximations.

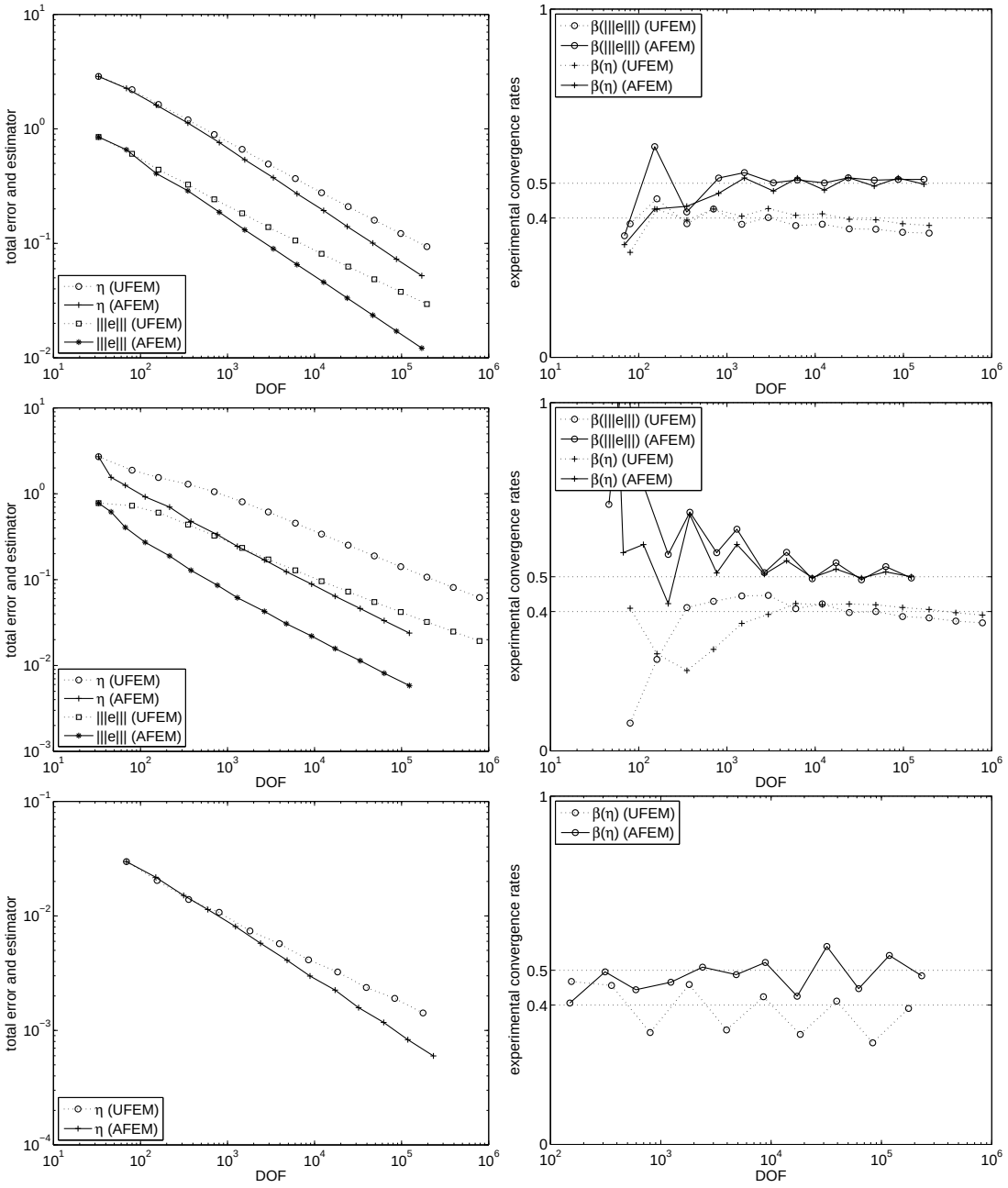


Figure 8.11: Example 1 (top), Example 2 (middle), Example 3 (bottom). Decrease of the estimator $\eta = \eta_\ell$ and the total error $|||e||| = |||e_\ell|||$ as a function of the DOF on a double logarithmic scale for uniform (UFEM) and adaptive (AFEM) refinement (left); experimental convergence rates of the estimator $\beta(\eta) = \beta_\ell(\eta_\ell)$ and of the total error $\beta(|||e|||) = \beta_\ell(|||e_\ell|||)$ as a function of the DOF on a semi-logarithmic scale (logarithmic abscissa, linear ordinate) for uniform (UFEM) and adaptive (AFEM) refinement (right). For Example 3 the quantities $|||e_\ell|||$ and $\beta_\ell(|||e_\ell|||)$ are not given since the exact solution is unknown.

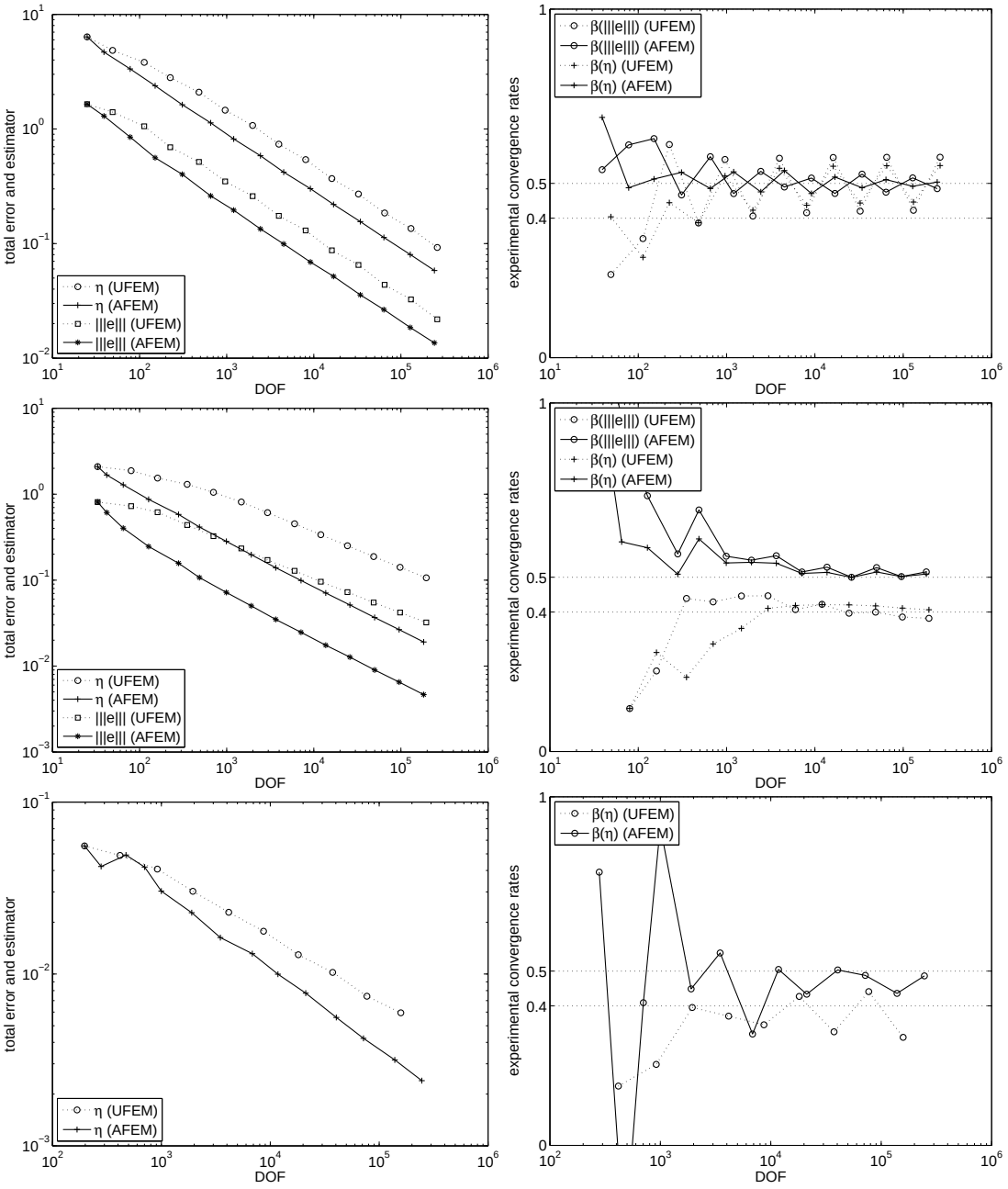


Figure 8.12: Example 4 (top), Example 5 (middle), Example 6 (bottom). Decrease of the estimator $\eta = \eta_\ell$ and the total error $|||e||| = |||e_\ell|||$ as a function of the DOF on a double logarithmic scale for uniform (UFEM) and adaptive (AFEM) refinement (left); experimental convergence rates of the estimator $\beta(\eta) = \beta_\ell(\eta_\ell)$ and of the total error $\beta(|||e|||) = \beta_\ell(|||e_\ell|||)$ as a function of the DOF on a semi-logarithmic scale for uniform (UFEM) and adaptive (AFEM) refinement (right). For Example 6 the quantities $|||e_\ell|||$ and $\beta_\ell(|||e_\ell|||)$ are not given since the exact solution is unknown.

8.4.3 Heuristic consistency error estimates

In the previous paragraph, we have seen that the adaptive algorithm based on the residual-type error estimator (6.1.1) allows to obtain convergence rates close to 0.5 for examples with less regular solutions. This result justifies the use of the adaptive algorithm 7.1 in general and of the residual a posteriori error estimates developed in this work in particular.

However, as we know from Paragraph 6.5.2, the estimate (6.1.1) involves two terms $e^c_{\ell,rel}$, $e^c_{\ell,eff}$ that are in general not computable. These terms, if not small in comparison to the error estimator η_ℓ, can drastically reduce the performance of the adaptive method, as the latter assumes that all the significant information about the error is contained in the computable part of the a posteriori estimate (cf. Assumption 6.5.1). Yet, there is no proof that Assumption 6.5.1 or any comparable result actually holds. To get out of this dilemma, computable heuristic estimates for the consistency errors were suggested in Section 6.6 (cf. Definition 6.6.11). In this paragraph, the related numerical results are documented and analyzed. Using the examples with known exact solution, the heuristic estimates are compared to the otherwise uncomputable exact consistency error quantities.

Heuristic estimates for the efficiency related consistency error

The decrease of the simplified form of the efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ (cf. (6.5.3)) and/or its estimate $E^c_{\ell,eff}$ as a function of the DOF is shown in Figures 8.13 and 8.14 for uniform and adaptive mesh refinement for all test examples. For Examples 1, 2, 4, and 5 where the analytical solution is known, the estimate $E^c_{\ell,eff}$ provides precise information about the magnitude and the decay rates of the exact consistency error $e^{c,L^2}_{\ell,eff}$. For the other examples where no exact solution is available, the quantities $E^c_{\ell,eff}$ predict a rapid reduction of the consistency error $e^{c,L^2}_{\ell,eff}$ term. The sharp estimation in case of the examples with known analytical solution, makes this prediction look credible.

Remark 8.4.3. The anomalous behavior of $e^{c,L^2}_{\ell,eff}$ for Example 5 in case of adaptive mesh refinement (Figure 8.14 middle right) is related to the quality of approximation of the discrete sets and will be addressed in the next paragraph. Note that even in this case the heuristic estimate $E^c_{\ell,eff}$ performs well.

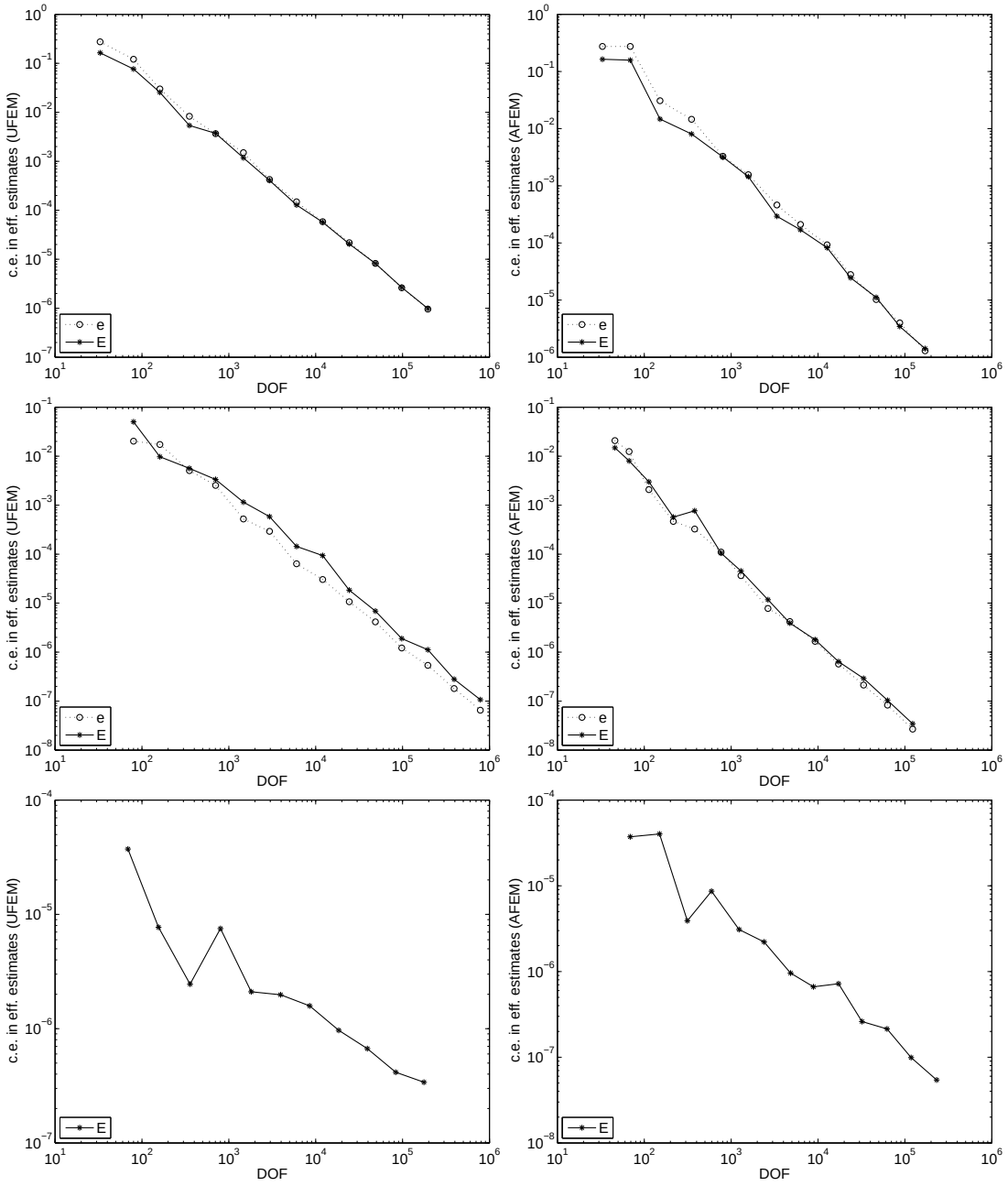


Figure 8.13: Example 1 (top), Example 2 (middle), Example 3 (bottom). Decrease of the efficiency related consistency error $e = e^{c,L^2}_{\ell,eff}$ and its estimate $E = E^{c}_{\ell,eff}$ as functions of the DOF on a double logarithmic scale for uniform refinement (left) and adaptive refinement (right). For Example 3 the quantity $e^{c,L^2}_{\ell,eff}$ is not given since the exact solution is unknown.

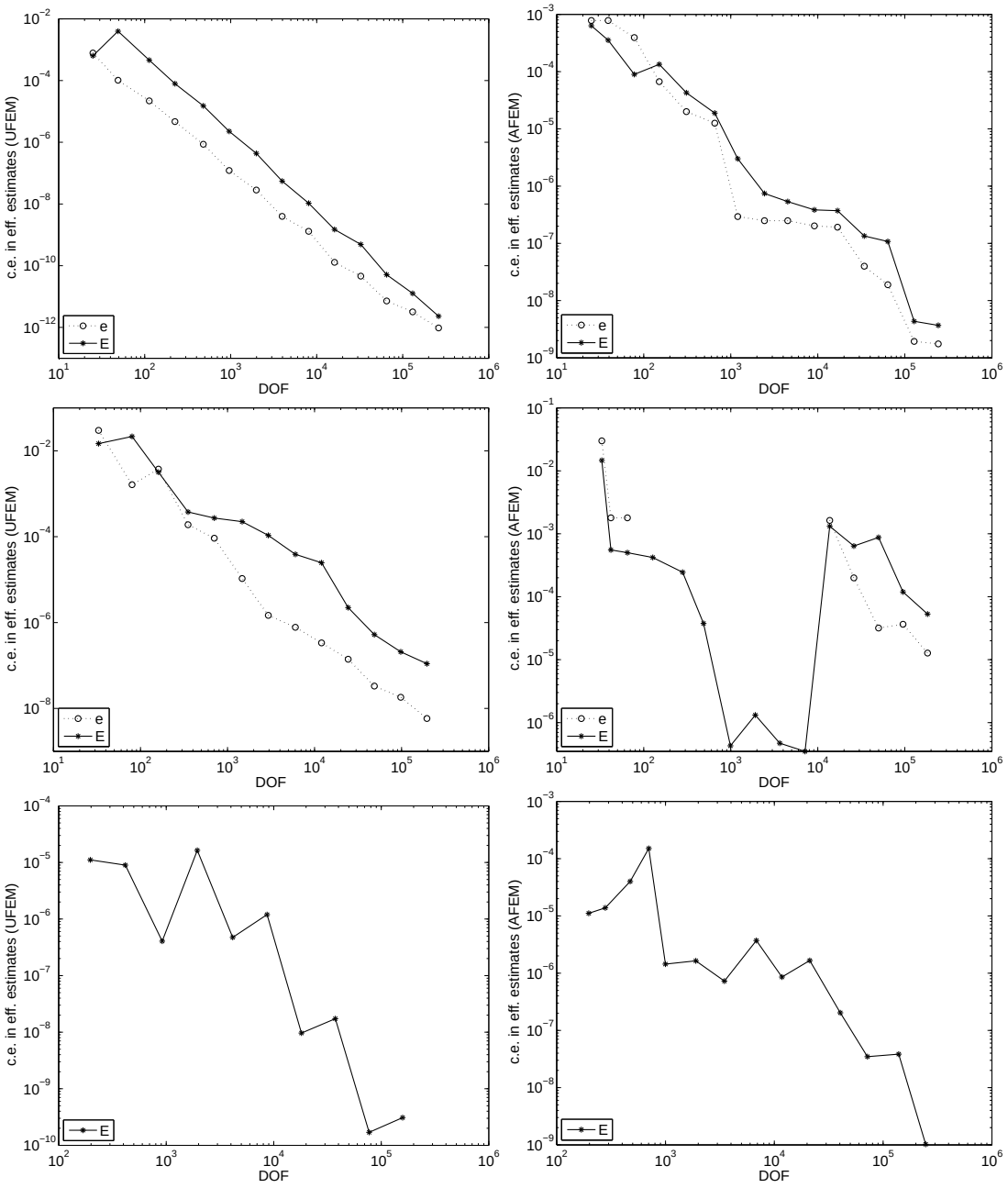


Figure 8.14: Example 4 (top), Example 5 (middle), Example 6 (bottom). Decrease of the efficiency related consistency error $e = e^{c,L^2}_{\ell,eff}$ and its estimate $E = E^c_{\ell,eff}$ as functions of the DOF on a double logarithmic scale for uniform refinement (left) and adaptive refinement (right). For Example 6 the quantity $e^{c,L^2}_{\ell,eff}$ is not given since the exact solution is unknown.

Heuristic estimates for the reliability related consistency error

Figures 8.15 and 8.16 document the decay of the absolute values of the reliability related consistency error $e^c_{\ell,rel}$ and/or of the associated heuristic estimates $E^{c,k}_{\ell,rel}$, $1 \leq k \leq 3$ as functions of DOF for uniform and adaptive mesh refinement for all test examples. In case of the reliability related consistency errors and their estimates, it is necessary to operate with the absolute values in order to enable a graphical representation of the results on a double logarithmic scale as the quantities $e^c_{\ell,rel}$ and $E^{c,k}_{\ell,rel}$, $1 \leq k \leq 3$ can be negative. Recall that $E^{c,1}_{\ell,rel}$ and $E^{c,2}_{\ell,rel}$ coincide for problems with strict complementarity (Examples 1, 2, and 3), therefore only $E^{c,2}_{\ell,rel}$ is shown in the respective plots.

Remark 8.4.4. Note that, in general, not the sign but the magnitude of the consistency errors is the crucial factor for the quality of the error estimate. Both, positive and negative consistency errors can affect the performance of the adaptive algorithm if they are not dominated by the error estimator.

The heuristic estimates $|E^{c,1}_{\ell,rel}|$ and/or $|E^{c,2}_{\ell,rel}|$ give proper upper bounds for the consistency error $|e^c_{\ell,rel}|$ for all examples with known solution. In case of Examples 1, 2, and 4 we see that $|E^{c,1}_{\ell,rel}|$ and/or $|E^{c,2}_{\ell,rel}|$ also provide approximately correct decay rates.

Figure 8.15 (middle) and Figure 8.16 (top) corresponding to Example 2 and 4, respectively, show that the heuristic estimate $|E^{c,3}_{\ell,rel}|$ underestimates the consistency error $|e^c_{\ell,rel}|$. Moreover, for the Examples 1, 2, 4, and 5, $E^{c,3}_{\ell,rel}$ often provides quantities with the wrong sign as it can be seen in the respective tables (e.g., see Tables A.1.6, A.1.7 for Example 1, and Tables A.2.6, A.2.7 for Example 2). The quantity $E^{c,3}_{\ell,rel}$ is therefore not suited for the estimation of the consistency error $e^c_{\ell,rel}$. The same imprecision regarding the sign is observed for $E^{c,2}_{\ell,rel}$ in case of Example 5 (see Tables A.5.7 and A.5.8). Therefore, we conclude that $|E^{c,2}_{\ell,rel}|$ should be used for problems with strict complementarity and $|E^{c,1}_{\ell,rel}|$ for problems with lack of strict complementarity.

In case of Example 5, the heuristic consistency error estimates of all three types are not sharp in the sense that they are grossly overestimating the consistency error with an insufficient decay rate. As a consequence, overall, the heuristic estimates only drop for around two orders of magnitude during the whole adaptive cycle.

In case of Examples 3 and 6, the exact consistency errors are not computable. For Example 3 (Figure 8.15, bottom), the quantity $|E^{c,2}_{\ell,rel}|$ predicts rapid decay rates of the reliability associated consistency error $|e^c_{\ell,rel}|$. In case of Example 6 (Figure 8.16, bottom), there is a similarity in the behavior of the decay rates of the heuristic estimates $|E^{c,1}_{\ell,rel}|$ and $|E^{c,2}_{\ell,rel}|$ with that observed for Example 5, namely, the heuristic estimates also only drop for around two orders of magnitude during the adaptive cycle. This similarity gives the impression that for Example 6, just as for Example 5, the heuristic estimates of the consistency error are for some reason not able to capture the decay rates of $|e^c_{\ell,rel}|$ correctly. Possible causes of the deficient performance of the heuristic estimates of the reliability related consistency errors in case of Example 5 and Example 6 will be discussed in the next paragraph.

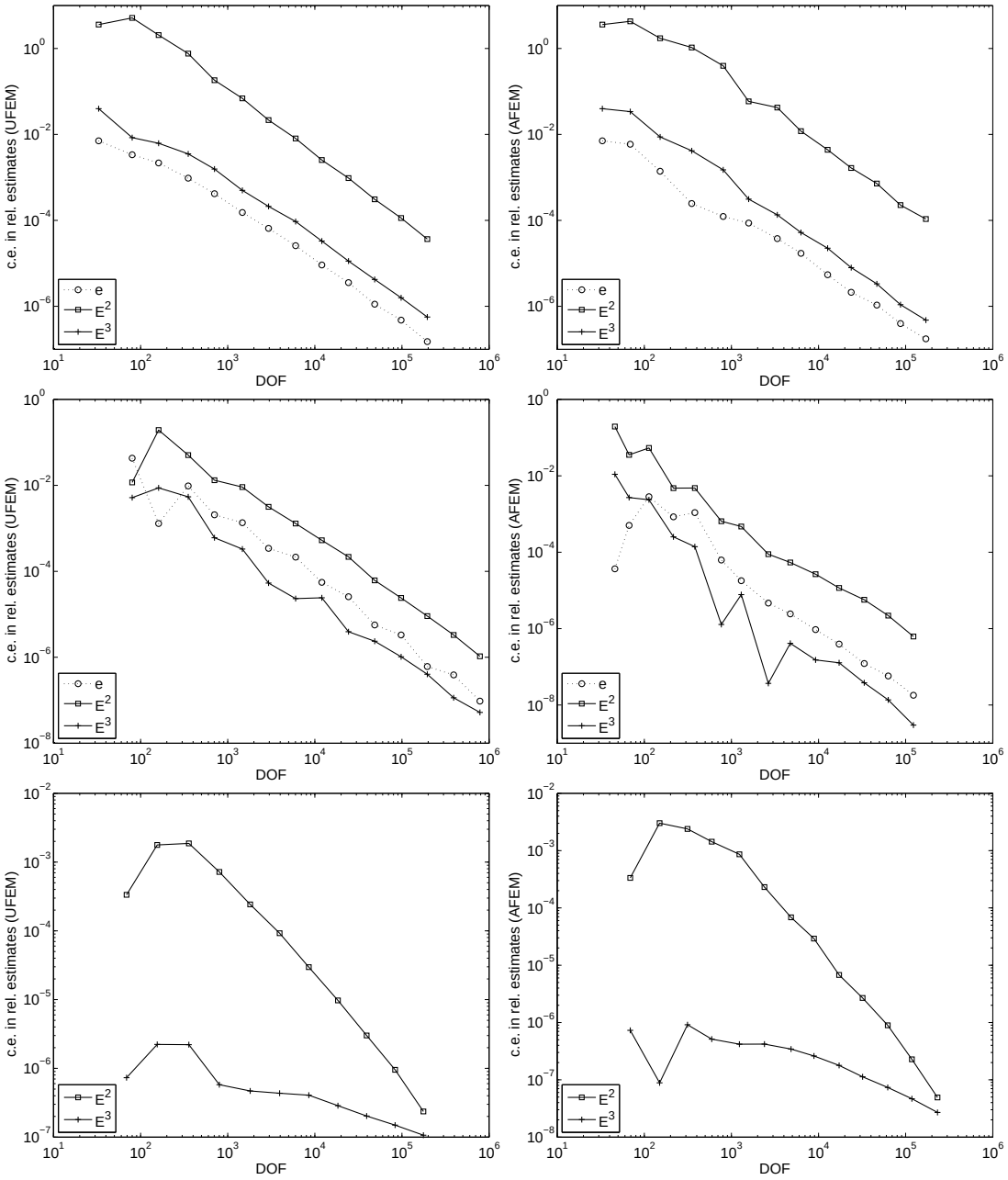


Figure 8.15: Example 1 (top), Example 2 (middle), Example 3 (bottom). Decrease of the absolute value of the reliability related consistency error $e = |e^c_{\ell,rel}|$ and its estimates $E^k = |E^{c,k}_{\ell,rel}|$, $k \in \{2,3\}$ as functions of the DOF on a double logarithmic scale for uniform refinement (left) and adaptive refinement (right). For Example 3 the quantity $|e^c_{\ell,rel}|$ is not given since the exact solution is unknown.

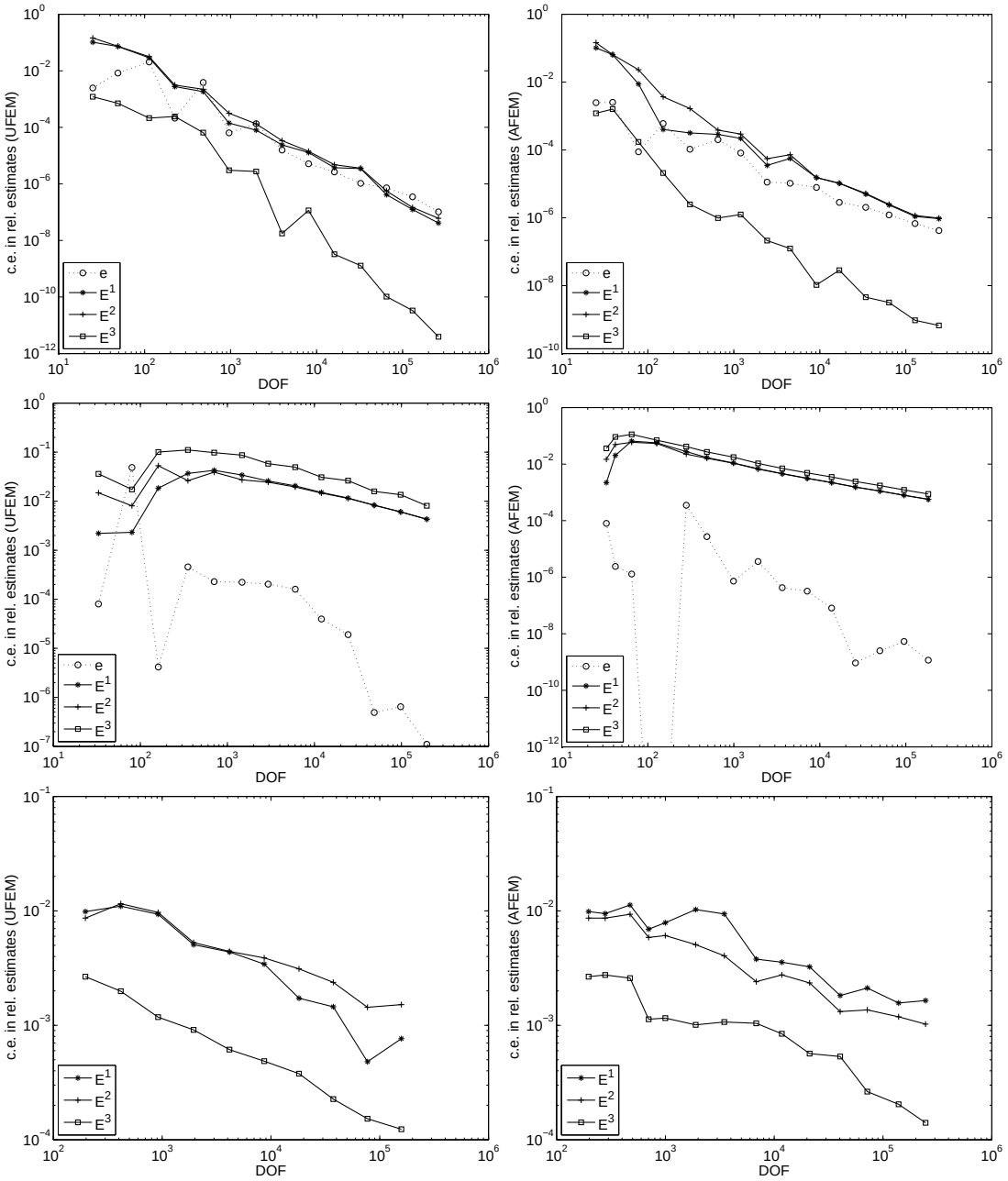


Figure 8.16: Example 4 (top), Example 5 ((middle), Example 6 (bottom). Decrease of the absolute value of the reliability related consistency error $e = |e^c_{\ell,rel}|$ and its estimates $E^k = |E^{c,k}_{\ell,rel}|$, $k \in \{1, 2, 3\}$ as functions of the DOF on a double logarithmic scale for uniform refinement (left) and adaptive refinement (right). For Example 6 the quantity $|e^c_{\ell,rel}|$ is not given since the exact solution is unknown.

8.4.4 Approximation error in the (strongly) active sets. Heuristic approximations of the exact sets

In this paragraph, we question the quality of the approximation of (strongly) active sets by the discrete solutions. In particular, it interesting to know whether Assumption 6.5.3 is satisfied in practice for the uniform and adaptive mesh refinement and if not, to investigate whether this fact influences the behavior of the exact and/or heuristically approximated consistency errors. Furthermore, for a better understanding of what might be the cause of the deficient performance of the heuristic estimates seen in Examples 5 and 6 in the previous paragraph, one can additionally look at the performance of the heuristic estimates of sets (cf. (6.6.3)), which form the basis of the heuristic estimates of the consistency errors.

We will measure the quality of the approximation of the continuous sets and the quality of the heuristic set approximations in terms of the quantities $e^{evd}_{\ell,\mathcal{A}}$, $e^{dva}_{\ell,\mathcal{A}}$, $e^{eva}_{\ell,\mathcal{A}}$ and $e^{evd}_{\ell,\mathcal{C}}$, $e^{dva}_{\ell,\mathcal{C}}$, $e^{eva}_{\ell,\mathcal{C}}$ (cf. (6.6.4)). Recall that these quantities correspond to the L^1-norm of the difference between the characteristic functions of the associated sets as illustrated in Figure 8.17. In particular, $e^{evd}_{\ell,\mathcal{A}}$ ($e^{evd}_{\ell,\mathcal{C}}$) shows how well the discrete active set $\mathcal{A}_\ell$ (discrete strongly active set $\mathcal{C}_\ell$) approximates the continuous active set $\mathcal{A}$ (strongly active set $\mathcal{C}$), whereas the quantities $e^{dva}_{\ell,\mathcal{A}}$, $e^{eva}_{\ell,\mathcal{A}}$ ($e^{dva}_{\ell,\mathcal{C}}$, $e^{eva}_{\ell,\mathcal{C}}$) reflect the quality of the heuristic approximation of the active set $\bar{\mathcal{A}}_\ell$ (of the strongly active set $\bar{\mathcal{C}}_\ell$).

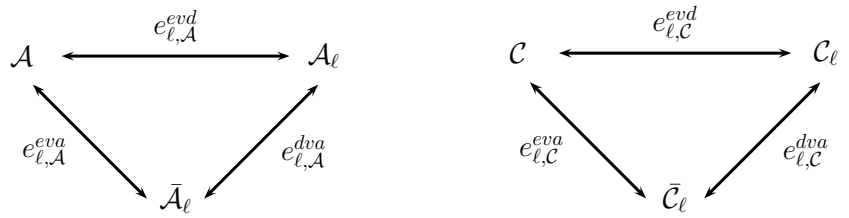


Figure 8.17: Scheme of error quantities associated with the active set $\mathcal{A}$ (left) and with the strongly active set $\mathcal{C}$ (right).

In terms of these three error quantities, it seems appropriate to say that the heuristic approximation of the active set works well if the following conditions are satisfied simultaneously:

$$e^{evd}_{\ell,\mathcal{A}} \approx e^{dva}_{\ell,\mathcal{A}} \quad \text{or at least} \quad e^{evd}_{\ell,\mathcal{A}} \leq e^{dva}_{\ell,\mathcal{A}} \text{ and } \beta_\ell(e^{evd}_{\ell,\mathcal{A}}) \approx \beta_\ell(e^{dva}_{\ell,\mathcal{A}})\,, \tag{8.4.1a}$$

$$e^{eva}_{\ell,\mathcal{A}} \ll e^{dva}_{\ell,\mathcal{A}}\,, \tag{8.4.1b}$$

$$e^{eva}_{\ell,\mathcal{A}} \ll e^{evd}_{\ell,\mathcal{A}}\,. \tag{8.4.1c}$$

And similarly, for the strongly active set

$$e^{evd}_{\ell,\mathcal{C}} \approx e^{dva}_{\ell,\mathcal{C}} \quad \text{or at least} \quad e^{evd}_{\ell,\mathcal{C}} \leq e^{dva}_{\ell,\mathcal{C}} \text{ and } \beta_\ell(e^{evd}_{\ell,\mathcal{C}}) \approx \beta_\ell(e^{dva}_{\ell,\mathcal{C}})\,, \tag{8.4.2a}$$

$$e^{eva}_{\ell,\mathcal{C}} \ll e^{dva}_{\ell,\mathcal{C}}\,, \tag{8.4.2b}$$

$$e^{eva}_{\ell,\mathcal{C}} \ll e^{evd}_{\ell,\mathcal{C}}\,. \tag{8.4.2c}$$

Test problems with strict complementarity

For Examples 1, 2, and 3, Figure 8.18 shows the decay of the error quantities $e^{dva}_{\ell,\mathcal{A}}$, $e^{evd}_{\ell,\mathcal{A}}$, and $e^{eva}_{\ell,\mathcal{A}}$ as functions on the DOF for the uniform and the adaptive mesh refinement. For Examples 1 and 2 all the quantities can be computed, whereas in case of Example 3 only $e^{dva}_{\ell,\mathcal{A}}$ is available.

Example 1 and Example 2 (with known analytical solution). We see that the continuous active sets $\mathcal{A}$ are well approximated by the discrete ones $\mathcal{A}_\ell$. Both adaptive and uniform refinement lead to a steady decay of the error $e^{evd}_{\ell,\mathcal{A}}$. In case of Example 1, uniform and adaptive algorithms lead to essentially the same results with respect to $e^{evd}_{\ell,\mathcal{A}}$, whereas in case of Example 2 the adaptive mesh refinement results in a slightly faster reduction of $e^{evd}_{\ell,\mathcal{A}}$. One can say that Assumption 6.5.3 is satisfied for these two examples.

Concerning the quality of the heuristic approximation of the active set $\bar{\mathcal{A}}_\ell$, we see that for both, the adaptive and uniform mesh refinement, the a posteriori quantity $e^{dva}_{\ell,\bar{\mathcal{A}}}$ yields an upper bound to the set error $e^{evd}_{\ell,\bar{\mathcal{A}}}$ with approximately the same decay rates, i.e., (8.4.1a) is satisfied.

We further examine the curves corresponding to $e^{eva}_{\ell,\bar{\mathcal{A}}}$, which measures the difference between the heuristic estimate of the active set $\bar{\mathcal{A}}_\ell$ and the exact active set $\mathcal{A}$ in comparison to that associated with $e^{evd}_{\ell,\bar{\mathcal{A}}}$ and its heuristic approximation $e^{dva}_{\ell,\bar{\mathcal{A}}}$. Obviously, $\bar{\mathcal{A}}_\ell$ does not provide a better approximation of the exact set $\mathcal{A}$ than the discrete active set $\mathcal{A}_\ell$ does. Namely, with essentially the same convergence rates, the three error quantities satisfy $e^{eva}_{\ell,\bar{\mathcal{A}}} \approx e^{dva}_{\ell,\bar{\mathcal{A}}}$, $e^{eva}_{\ell,\bar{\mathcal{A}}} \geq e^{evd}_{\ell,\bar{\mathcal{A}}}$ for Example 1, whereas for Example 2, $e^{eva}_{\ell,\bar{\mathcal{A}}} \leq e^{dva}_{\ell,\bar{\mathcal{A}}}$, $e^{eva}_{\ell,\bar{\mathcal{A}}} \approx e^{evd}_{\ell,\bar{\mathcal{A}}}$ in case of uniform mesh refinement and $e^{eva}_{\ell,\bar{\mathcal{A}}} \leq e^{dva}_{\ell,\bar{\mathcal{A}}}$, $e^{eva}_{\ell,\bar{\mathcal{A}}} \geq e^{evd}_{\ell,\bar{\mathcal{A}}}$ in case of adaptive mesh refinement. This behavior does not align with the ideal expectation (8.4.1b) and (8.4.1c). The fact that $e^{eva}_{\ell,\bar{\mathcal{A}}}$ is not much smaller than $e^{dva}_{\ell,\bar{\mathcal{A}}}$ explains our observation from the previous paragraph that the heuristic estimate $E^{c,2}_{\ell,rel}$ of the reliability associated consistency error is not sharp with respect to the exact consistency error $e^{c}_{\ell,rel}$ (see Figure 8.15, top and middle), although, it gives a proper upper bound with a correct decay rate.

Example 3 (no analytical solution). Relying on the information provided by $e^{dva}_{\ell,\bar{\mathcal{A}}}$ in case of Example 3, one can say that adaptive refinement improves the quality of the approximation of the active set $\mathcal{A}$, since $e^{dva}_{\ell,\bar{\mathcal{A}}}$ reaches lower values.

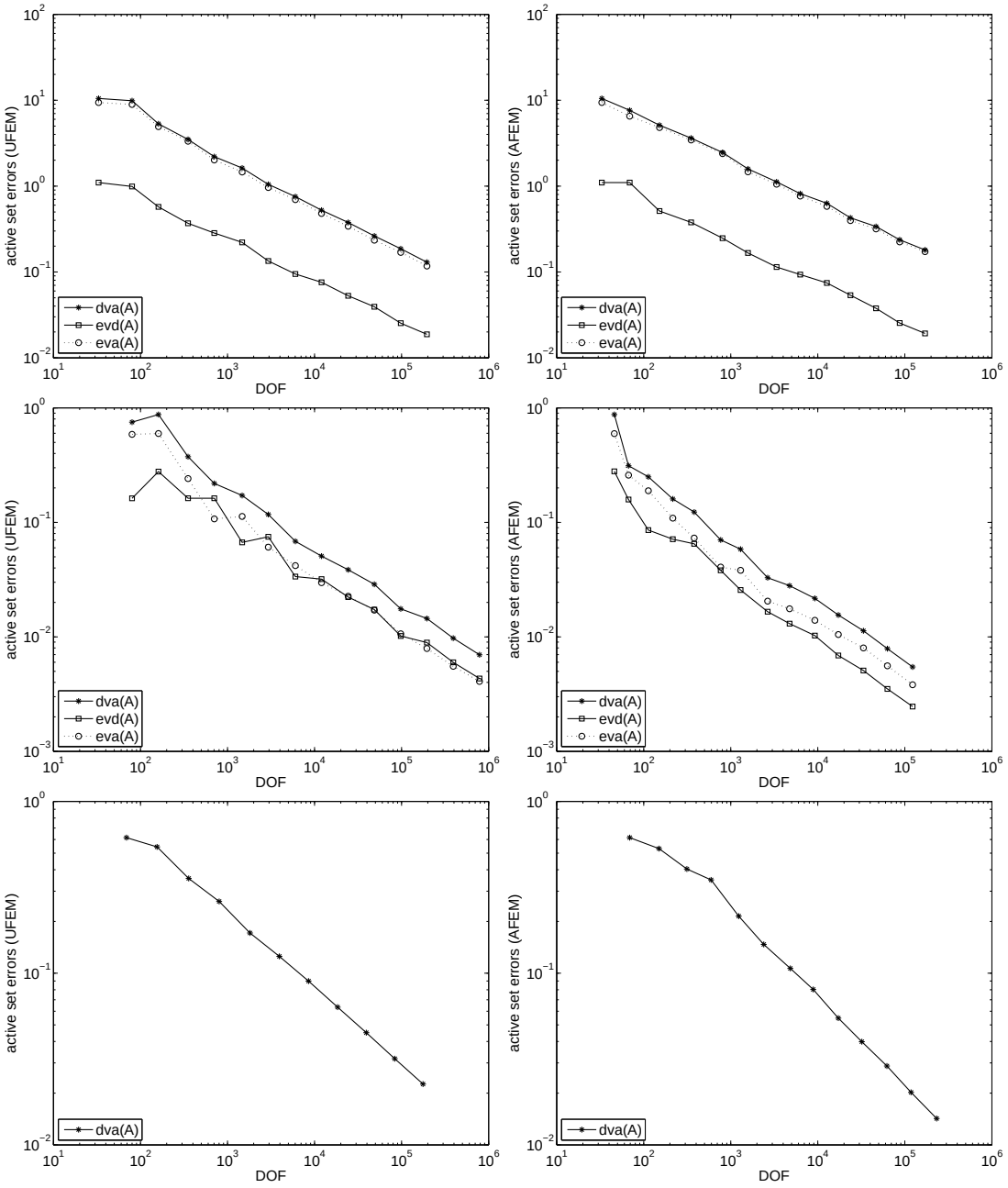


Figure 8.18: Approximation of the active set. Example 1 (top), Example 2 (middle), Example 3 (bottom). Quantities $dva(A) = e^{dva}_{\ell,\mathcal{A}}$, $evd(A) = e^{evd}_{\ell,\mathcal{A}}$, and $eva(A) = e^{eva}_{\ell,\mathcal{A}}$ as functions of the DOF on a double logarithmic scale for uniform refinement (left) and adaptive refinement (right). For Example 3 the quantities $e^{evd}_{\ell,\mathcal{A}}$ and $e^{eva}_{\ell,\mathcal{A}}$ are not given since the exact solution is unknown.

Test problems with lack of strict complementarity

For Examples 4, 5, and 6, Figures 8.19 and 8.20 show the decay of the errors quantities $e^{dva}_{\ell,\mathcal{A}}$, $e^{evd}_{\ell,\mathcal{A}}$, $e^{eva}_{\ell,\mathcal{A}}$ and $e^{dva}_{\ell,\mathcal{C}}$, $e^{evd}_{\ell,\mathcal{C}}$, $e^{eva}_{\ell,\mathcal{C}}$, respectively, as functions of the DOF for the uniform and the adaptive mesh refinement.

It should be emphasized that, in case of the examples with lack of strict complementarity, it is a priori clear that the results with respect to the convergence of the discrete active and strongly active sets in terms of quantities $e^{evd}_{\ell,\mathcal{A}}$, $e^{evd}_{\ell,\mathcal{C}}$ cannot be expected to be as satisfactory as for the problems with strict complementarity. As is was observed in Paragraph 8.4.1, the strongly active set is not well resolved by the adaptive algorithm. Moreover, in the algorithm PDASS(LSC,ϵ) the discrete active sets $\mathcal{A}_\ell$ are artificially constructed using the parameter ϵ whose value is selected based on a heuristical assumption (cf. Section 8.2). This fact influences the convergence of the discrete active sets $\mathcal{A}_\ell$ and can be observed in the behavior of the quantity $e^{evd}_{\ell,\mathcal{A}}$.

Example 4 and Example 5 (with known analytical solution). In case of Example 4 (Figure 8.19 top), after a quick decay at the first several iterations of the uniform algorithm (left) and the adaptive algorithm (right), the approximation error associated with the active set $e^{evd}_{\ell,\mathcal{A}}$ starts to increase. For Example 5 (Figure 8.19 middle), in case of the uniform refinement, $e^{evd}_{\ell,\mathcal{A}}$ experiences significant rises and drops in the magnitude as the number of DOF increases, at the last iteration finally achieving the value slightly below 10^{-2}. In case of adaptive mesh refinement, $e^{evd}_{\ell,\mathcal{A}}$ decreases with gradually diminishing slope, at $DOF \approx 10^4$ stabilizing at the value slightly under 10^{-2}. Presumably, this behavior is due to the artificial detection of the discrete active set $\mathcal{A}_\ell$ in PDASS(LSC,ϵ).

Unlike the discrete active set $\mathcal{A}_\ell$, the discrete strongly active set $\mathcal{C}_\ell$ is provided by the PDASS(LSC,ϵ) algorithm directly. In case of the uniform mesh refinement, we therefore observe better convergence results for $\mathcal{C}_\ell$. In case of Example 4 (Figure 8.20 top left), $e^{evd}_{\ell,\mathcal{C}}$ is decreasing with essentially constant decay rate. For Example 5 (Figure 8.20 middle left), after achieving $DOF > 10^3$ necessary for a proper resolution of $\mathcal{C}_\ell$, the value of $e^{evd}_{\ell,\mathcal{C}}$ decreases with a stable decay rate.

In case of the adaptive mesh refinement, convergence results related to $\mathcal{C}$ are less optimistic then for the uniformly mesh refinement. This is due to the poor resolution of the mesh around the strongly active set provided by the adaptive algorithm (cf. Paragraph 8.4.1). In case of Example 4 where we have $N^C_\ell \leq 66$, $0 \leq \ell \leq 14$ (see Table A.4.12), we observe an effect that sometimes occurs on such rough meshes: an addition of a couple of DOF at each subsequent iteration of the adaptive loop leads to a gradual increase of the approximation error $e^{evd}_{\ell,\mathcal{C}}$ at the second half of the adaptive cycle. In case of Example 5 (Figure 8.20 middle right), the resolution of the strongly active set stays practically unchanged until the overall number of DOF approaches the value 10^4, that is when the value of $e^{evd}_{\ell,\mathcal{C}}$ starts to decrease. At the last iteration, however, the resolution of $\mathcal{C}$ is still very poor so that $e^{evd}_{\ell,\mathcal{C}} < 10^{-1}$. Note that at the last iteration of the adaptive loop $N^C_\ell = 245$ (cf. Table A.5.12). This behavior of the discrete strongly active set allows to explain the sudden rise of the value of the efficiency related consistency error

$e^{c,L^2}_{\ell,eff}$ that can be seen in Figure 8.14 middle right: with $\mu = 0$ in $\mathcal{B}$ (see the description of Example 5) and the discrete inactive sets $\mathcal{I}_\ell$ satisfying $\mathcal{I}_\ell \subset \mathcal{I} \cup \mathcal{B}$, the expression of $e^{c,L^2}_{\ell,eff}$ reduces to

$$e^{c,L^2}_{\ell,eff} = \sum_{T \subseteq \mathcal{Z}_\ell} h_T^2 \|\sigma\|^2_{0,T}$$

(cf. (6.5.3)). In the significant part of the adaptive cycle ($3 \leq \ell \leq 9$), the discrete strongly active set $\mathcal{C}_\ell$ stays almost unchanged (see Figure 8.20, middle right). The curve corresponding to $e^{c,L^2}_{\ell,eff}$ in Figure 8.14 middle right tells us that for $3 \leq \ell \leq 9$, the set $\mathcal{C}_\ell$ lies strictly within the exact set $\mathcal{C}$, so that $\mathcal{C} \cap \mathcal{Z}_\ell = \varnothing$ holds true. Therefore, $e^{c,L^2}_{\ell,eff} = 0$, $3 \leq \ell \leq 9$ and the corresponding data points are omitted in the plot.

Concerning the quality of the heuristic approximation of the continuous active and strongly active sets, in case of Example 5, the quantities $e^{evd}_{\ell,\mathcal{A}}$, $e^{eva}_{\ell,\mathcal{A}}$, as well as $e^{evd}_{\ell,\mathcal{C}}$, $e^{eva}_{\ell,\mathcal{C}}$ show approximately the same behavior as in case of the examples with strict complementarity. That is, $e^{dva}_{\ell,\mathcal{A}}$ ($e^{dva}_{\ell,\mathcal{C}}$) provides an upper bound to $e^{evd}_{\ell,\mathcal{A}}$ ($e^{evd}_{\ell,\mathcal{C}}$) with similar convergence rates, whereas the magnitude of $e^{eva}_{\ell,\mathcal{A}}$ ($e^{eva}_{\ell,\mathcal{C}}$) is not much smaller comparing to that of the other two quantities. In case of Example 4, the heuristic estimates of the (strongly) active set perform badly except in case of strongly active set and uniform refinement that features $e^{evd}_{\ell,\mathcal{C}} \approx e^{dva}_{\ell,\mathcal{C}} \approx e^{eva}_{\ell,\mathcal{C}}$.

Remark 8.4.5. By varying the parameters γ and r in the definition of the approximations of the characteristic functions (6.6.1), one could possibly construct heuristic approximations for the (strongly) active sets that satisfy (8.4.1) ((8.4.2)) more closely.

Example 6 (no analytical solution). The heuristic approximation of the discretization error in the active set $e^{dva}_{\ell,\mathcal{A}}$ stops decreasing after reaching the value $e^{dva}_{\ell,\mathcal{A}} \approx 10^{-2}$. This behavior is similar to that observed in Examples 4 and 5 and is presumably caused by the imprecision in detection of the discrete active set $\mathcal{A}_\ell$. The error quantity $e^{dva}_{\ell,\mathcal{C}}$ in case of uniform mesh refinement behaves unstably for this example due to chattering observed at each iteration of the PDASSS(LSC,ϵ) algorithm (cf. Remark 8.4.2). In case of the adaptive mesh refinement, chattering did not occur at some iterations of the adaptive loop. As we can see, as soon as a sufficient resolution of the strongly active set is achieved, $e^{dva}_{\ell,\mathcal{C}}$ decreases with an essentially unchanging convergence rate, not influenced by the chattering of the solver as much as in case of the uniform mesh refinement.

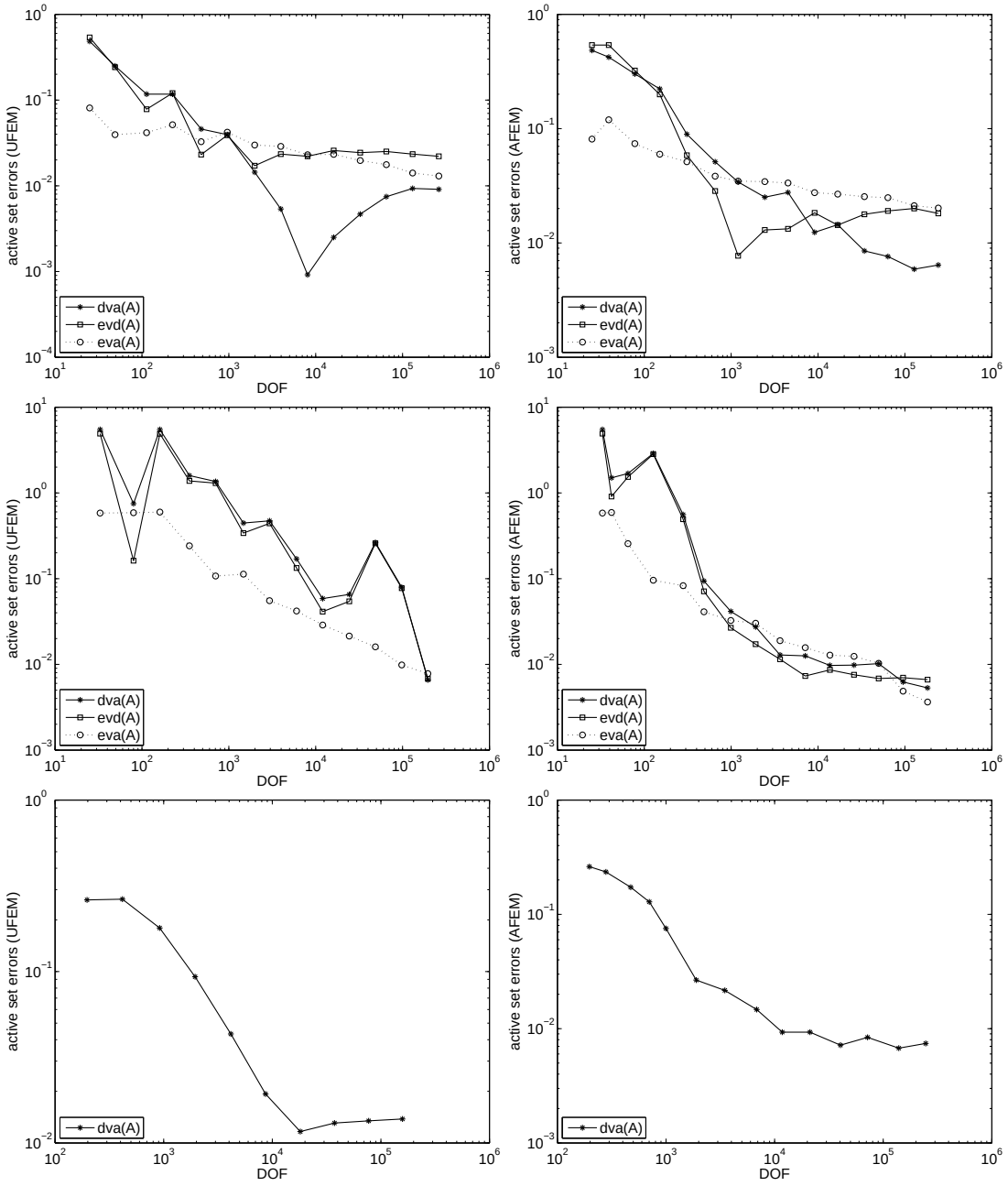


Figure 8.19: Approximation of the active set. Example 4 (top), Example 5 (middle), Example 6 (bottom). Quantities $dva(A) = e^{dva}_{\ell,\mathcal{A}}$, $evd(A) = e^{evd}_{\ell,\mathcal{A}}$, and $eva(A) = e^{eva}_{\ell,\mathcal{A}}$ as functions of the DOF on a double logarithmic scale for uniform refinement (left) and adaptive refinement (right). For Example 6 the quantities $e^{evd}_{\ell,\mathcal{A}}$ and $e^{eva}_{\ell,\mathcal{A}}$ are not given since the exact solution is unknown.

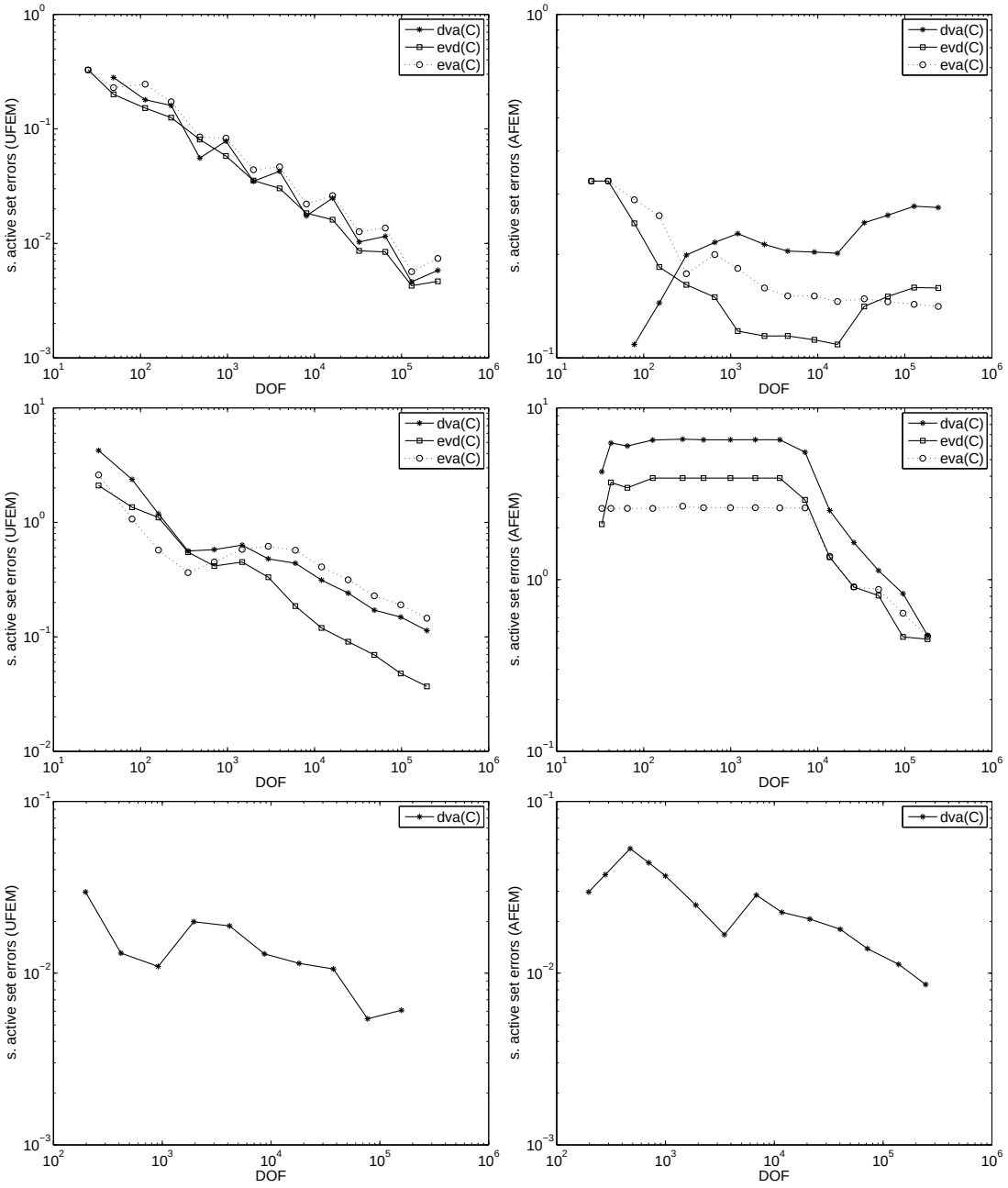


Figure 8.20: Approximation of the strongly active set. Example 4 (top), Example 5 (middle), Example 6 (bottom). Quantities $dva(C) = e^{dva}_{\ell,\mathcal{C}}$, $evd(C) = e^{evd}_{\ell,\mathcal{C}}$, and $eva(C) = e^{eva}_{\ell,\mathcal{C}}$ as functions of the DOF on a double logarithmic scale for uniform refinement (left) and adaptive refinement (right). For Example 6 the quantities $e^{evd}_{\ell,\mathcal{C}}$ and $e^{eva}_{\ell,\mathcal{C}}$ are not given since the exact solution is unknown.

Performance of the heuristic consistency error estimates of $e^c_{\ell,rel}$. Synopsis

In Paragraph 8.4.3, we have seen that the heuristic estimates of $e^c_{\ell,rel}$ perform badly for Example 5. They provide upper bounds but are not able to capture the decay rate of the exact consistency error $e^c_{\ell,rel}$ correctly. It should be noted that the structure of Example 5 is very similar to that of Example 2. The substantial difference lies in the fact that Example 2 satisfies the strict complementarity condition. The heuristic consistency error estimate $E^{c,2}_{\ell,rel}$ for Example 2 provides both, a valid upper bound and correct information with regard to the decay rates. The analysis of the heuristic set approximations $\bar{\mathcal{A}}_\ell$ and $\bar{\mathcal{C}}_\ell$ in terms of the error quantities $e^{evd}_{\ell,\mathcal{A}}$, $e^{dva}_{\ell,\mathcal{A}}$, $e^{eva}_{\ell,\mathcal{A}}$ and $e^{evd}_{\ell,\mathcal{C}}$, $e^{dva}_{\ell,\mathcal{C}}$, $e^{eva}_{\ell,\mathcal{C}}$ that have been performed in this paragraph, shows no significant differences in the performance of the heuristic set estimates for Example 2 and Example 5. In both cases, the quantities $e^{dva}_{\ell,\mathcal{A}}$ and $e^{dva}_{\ell,\mathcal{C}}$ give correct information about the behavior of $e^{evd}_{\ell,\mathcal{A}}$ and $e^{evd}_{\ell,\mathcal{C}}$, respectively. We can therefore exclude that the heuristic set estimates are responsible for the deficient performance of the reliability associated heuristic consistency error estimates. However, as we know, in case of Example 5 other effects specific for the numerical approximations of problems with lack of strict complementarity occur. These are:

(i) a deficient approximation of the active set due to the artificial detection with a heuristically chosen parameter ϵ employed in PDASS(LSC,ϵ);

(ii) poor resolution of the strongly active set in case of the adaptive mesh refinement.

It seems that the effect of factor (ii) should be less significant than the effect of factor (i): in this paragraph we have seen that the strongly active set is well resolved in case of the uniform mesh refinement (see Figure 8.21 middle). Nevertheless, the results presented in Paragraph 8.4.3 show that the heuristic estimates $E^{c,1}_{\ell,rel}$ and $E^{c,2}_{\ell,rel}$ are similarly imprecise for both, the uniform and adaptive mesh refinement (see Figure 8.16 middle).

Remark 8.4.6. It is, of course, not excluded that a further detailed investigation of the methods of a posteriori error control for the optimally controlled obstacle problem will reveal an improvement with regard to factor (ii) to be necessary. Therefore, it should be noted that the adaptive algorithm 7.1 intrinsically ignores the strongly active set. Being based on the a posteriori error estimate (6.1.1), its goal is to minimize the total error $|||e_\ell|||^2 = \|e_{\ell,y}\|^2_{1,\Omega} + \|e_{\ell,p}\|^2_{1,\Omega} + \|e_{\ell,u}\|^2_{0,\Omega}$ that does not contain any direct information about the error in the multiplier σ whose shape determines the structure of the strongly active set. In order to improve factor (ii), one has to choose a different measure of the approximation error in the a posteriori error analysis.

The heuristic set approximations and consistency error estimates are constructed from the discrete solutions y_ℓ, p_ℓ, σ_ℓ and the associated discrete sets $\mathcal{A}_\ell$ and $\mathcal{C}_\ell$. Thus, the imprecision in the discrete solution caused by factor (i) can have a bad influence on the performance of the heuristic estimates. For Example 6 presumably the same effect

occurs, worsened by the chattering of the solver. Factor (i) as well as the chattering phenomenon can be avoided by employing different kinds of solvers in step SOLVE (cf. Remark 7.2.5). This modification alone could presumably improve the performance of the heuristic error estimates associated with the reliability estimate.

Remark 8.4.7. It is interesting that under the same circumstances regarding factors (i) and (ii), the heuristic consistency error estimates work well in case of Example 4 and significantly overestimate in case of Example 5. This effect could be caused by the difference in the structure of the solutions of these two problems: in the vicinity of the free boundary $\mathcal{F}(y)$ the (adjoint) state functions change much quicker in case of Example 5, therefore, the effect of factor (i) is enhanced.

Remark 8.4.8. Remarkably, the heuristic estimate $E^c_{\ell,eff}$ performs equally well for all test problems, with and without strict complementarity. This has to do with the structure of the estimate $E^c_{\ell,eff}$. Unlike $E^{c,1}_{\ell,rel}$ and $E^{c,2}_{\ell,rel}$, its expression does not involve the data functions (cf. (6.6.7), (6.6.9), and (6.6.10)). Thus, the effects described above do not occur in case of $E^c_{\ell,eff}$. Unfortunately, as we have seen in Paragraph 8.4.3, a similar construction for the reliability related consistency error $E^{c,3}_{\ell,rel}$ does not provide a proper upper bound to $e^c_{\ell,rel}$.

8.4.5 Components of the a posteriori error estimate

In this paragraph, the results of the numerical experiments are examined from the point of view of Assumption 6.5.1, that is, the magnitudes of the error estimator η^2_ℓ are compared with the magnitudes of the exact consistency errors $e^{c,L^2}_{\ell,eff}$ and $|e^c_{\ell,rel}|$. Additionally, the heuristic estimates $E^c_{\ell,eff}$, as well as $|E^{c,2}_{\ell,rel}|$ in case of examples with strict complementarity and $|E^{c,1}_{\ell,rel}|$ for examples with lack of strict complementarity are included into all plots in order to see what prognosis the heuristic consistency error estimates give on the ratios (6.5.5). Note that for problems without analytical solution, the heuristic estimates are the only source of information with regard to the uncomputable consistency error terms. In the end of the paragraph, the magnitude of the data oscillation terms is also compared to that of η^2_ℓ.

Figures 8.21 and 8.22 show the decrease of the estimator η^2_ℓ, the absolute value of the reliability related consistency error $|e^c_{\ell,rel}|$ with its heuristic estimate $|E^{c,2}_{\ell,rel}|$, and the efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ with its heuristic estimate $E^c_{\ell,eff}$ as functions of the DOF for uniform refinement and adaptive mesh refinement.

For Examples 1, 2, 4, and 5 the exct consistency errors $e^{c,L^2}_{\ell,eff}$ and $|e^c_{\ell,rel}|$ are significantly dominated by the error estimator η^2_ℓ. Moreover, the decay rates of $e^{c,L^2}_{\ell,eff}$ and $|e^c_{\ell,rel}|$ are at least as high as that of η^2_ℓ. The only exception occurs in case of Example 5 for adaptive mesh refinement (see Figure 8.22 middle right). Overall, η^2_ℓ dominates $e^{c,L^2}_{\ell,eff}$. However, due to the extremely poor resolution of the strongly active set, the consistency error $e^{c,L^2}_{\ell,eff}$, being equal to zero at iterations $3 \leq \ell \leq 9$, acquires non-zero

values only at iterations $10 \leq \ell \leq 13$ (see the discussion in Paragraph 8.4.4), at iteration $\ell = 10$ getting close to (but staying below) the value of η_ℓ^2 . Nevertheless, it then rapidly decreases and gains a difference to η_ℓ^2 of almost two orders of magnitude.

In case of Examples 1, 2, and 4, the heuristic error estimates correctly predict that the consistency terms decay at least as fast as the error estimator. In case of Example 1, $|E_{\ell,rel}^{c,2}|$ grossly overestimates $|e_{\ell,rel}^c|$ (see Figure 8.21 top right), nevertheless, the heuristic estimates are significantly smaller then the error estimator.

In case of Example 5, due to the significant overestimation with insufficient decay rates, the heuristic estimate $|E_{\ell,rel}^{c,1}|$ gives false information about the ratio between the reliability related consistency error and the error estimator whereas $E_{\ell,eff}^c$ predicts the behavior of η_ℓ^2 essentially correct (Figure 8.22, middle).

In case of Examples 3 and 6, the exact consistency errors are not known. Due to the chattering of the solver, strong overestimation of the exact consistency errors can be expected. Indeed, for Example 3 the heuristic estimate $|E_{\ell,rel}^{c,2}|$ exceeds η_ℓ^2 for DOF below $2 * 10^4$. Still, $|E_{\ell,rel}^{c,2}|$ shows that the consistency error decreases much faster than η_ℓ^2. Furthermore, $E_{\ell,eff}^c$ is dominated by η_ℓ^2. Thus, overall, one can say that in this example the consistency errors should not influence the performance of the adaptive mesh refinement based on η_ℓ.

As we have seen in Paragraphs 8.4.3 and 8.4.4, presumably the same deficient performance of $|E_{\ell,rel}^{c,1}|$ observed for Example 5 is also relevant in case of Example 6, worsened by the chattering of the solver. Whereas the efficiency related heuristic estimate $E_{\ell,eff}^c$ is significantly dominated by the estimator, $|E_{\ell,rel}^{c,1}|$ dramatically dominates η_ℓ^2 for both, the uniform and adaptive mesh refinement.

Figures 8.23 and 8.24 show the decrease of the estimator η_ℓ^2 and the efficiency related data oscillation term $osc_{\ell,eff}^2$ as functions of the DOF for uniform and adaptive mesh refinement. For Example 6, $osc_{\ell,rel} = osc_{\ell,eff} = 0$. For all the other examples, $osc_{\ell,rel} = 0$ and the data oscillation $osc_{\ell,eff}$ is decreasing much faster and is significantly dominated by the error estimator.

Remark 8.4.9. As we have seen in Paragraph 8.4.4, for problems with strict complementarity, Assumption 6.5.3 is satisfied for both, the uniform and the adaptive mesh refinement. For problems with lack of strict complementarity on the other hand, Assumption 6.5.3 cannot in general be fulfilled with respect to the zero set $\mathcal{Z}$ in case of the adaptive mesh refinement (see Remark 8.4.6), even if the quality of the approximation of $\mathcal{A}$ is improved by employing a different algorithm in step SOLVE. However, as Figure 8.22 shows, the exact consistency errors corresponding to Examples 4 and 5 in case of adaptive mesh refinement are dominated by the error estimator η_ℓ^2 and, moreover, decay at least as fast as η_ℓ^2. These results give the impression that part of Assumption 6.5.3 with respect to the zero set might be redundant.

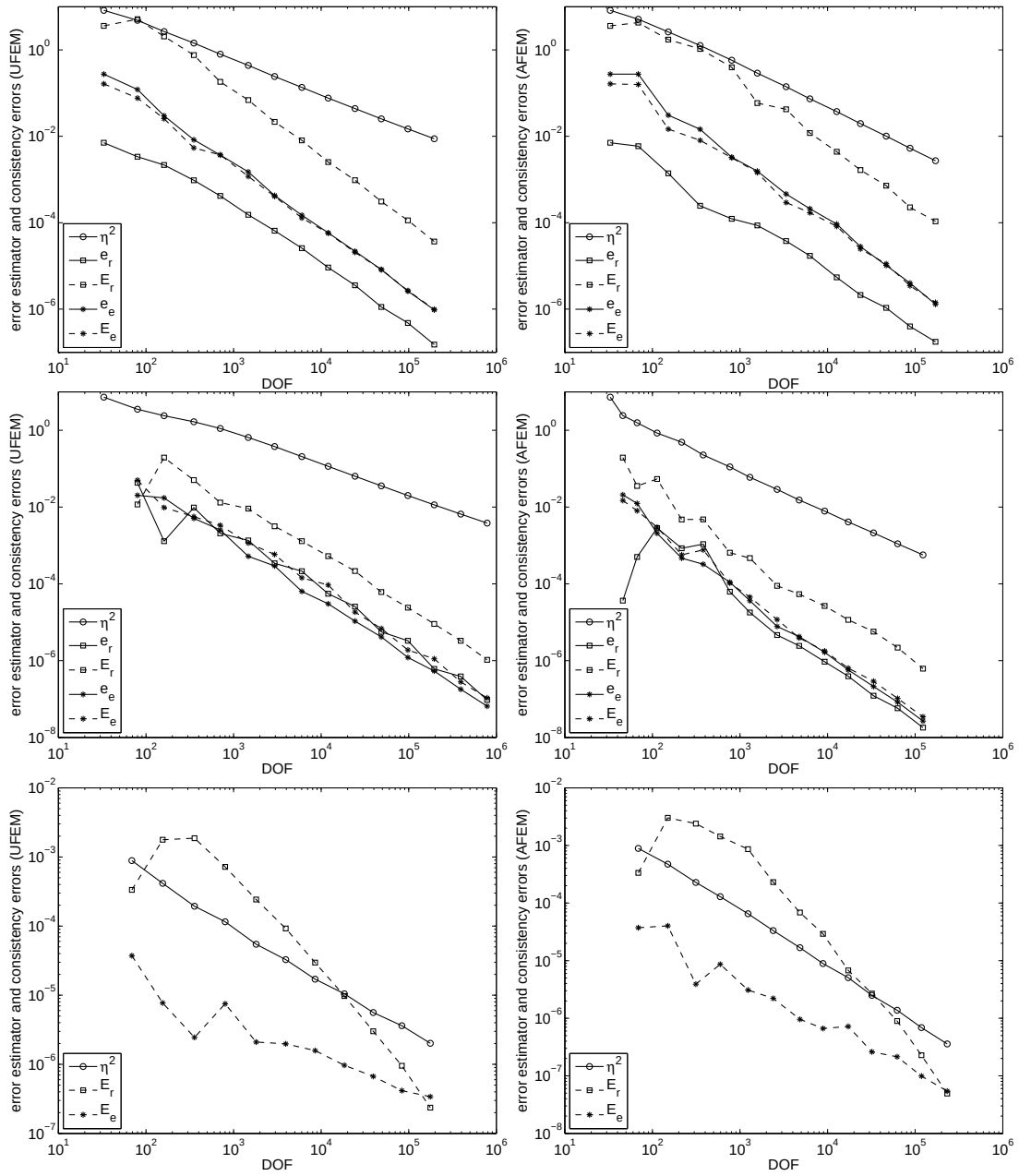


Figure 8.21: Example 1 (top), Example 2 (middle), Example 3 (bottom). Decrease of the estimator $\eta^2 = \eta_\ell^2$, the absolute value of the reliability related consistency error $e_r = |e^c_{\ell,rel}|$ and its heuristic estimate $E_r = |E^{c,2}_{\ell,rel}|$, and the efficiency related consistency error $e_e = e^{c,L^2}_{\ell,eff}$ and its heuristic estimate $E_e = E^c_{\ell,eff}$ as functions of the DOF on a double logarithmic scale for uniform refinement (left) and adaptive refinement (right). For Example 3 the quantities $|e^c_{\ell,rel}|$ and $e^{c,L^2}_{\ell,eff}$ are not given since the exact solution is unknown.

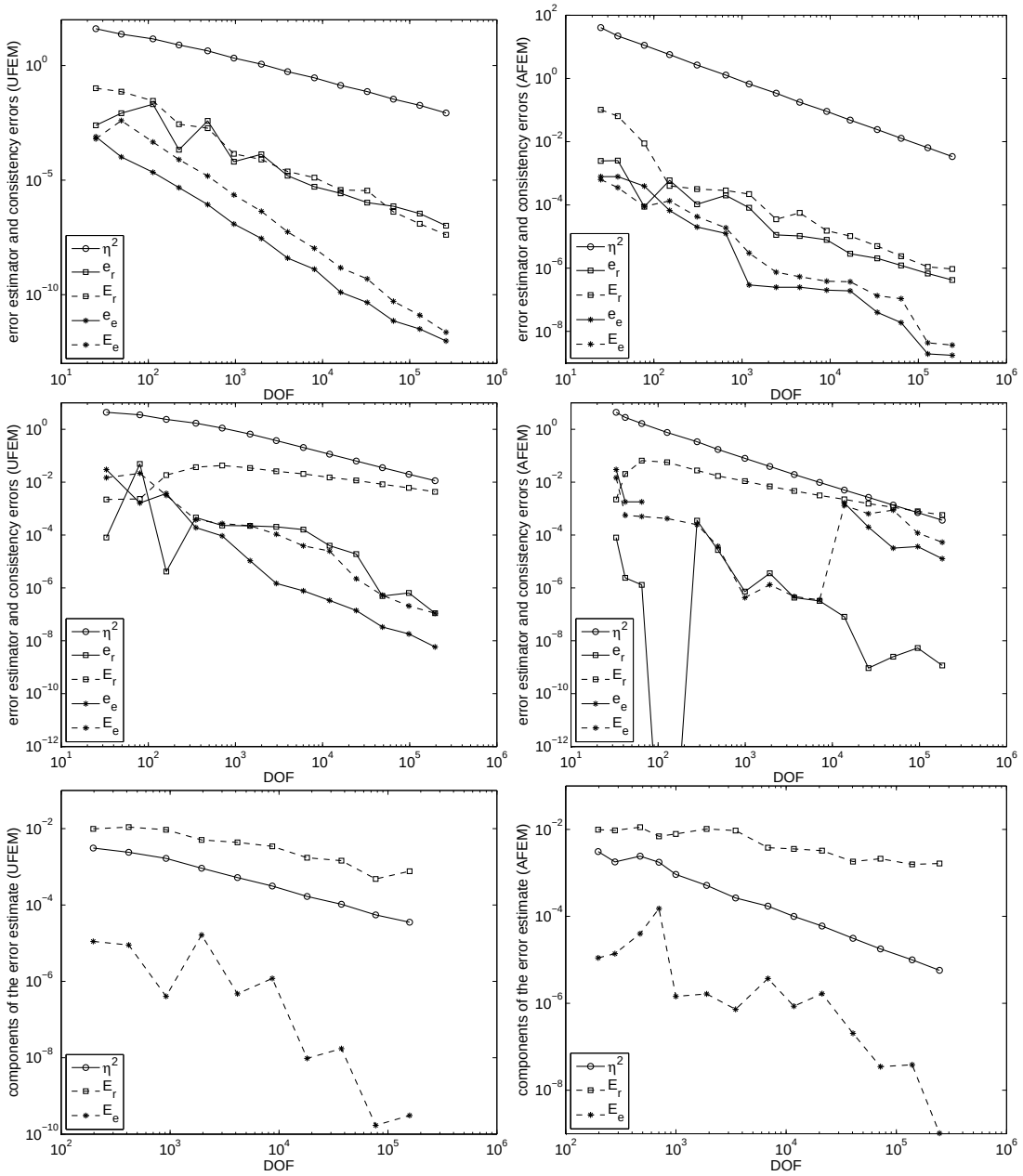


Figure 8.22: Example 4 (top), Example 5 (middle), Example 6 (bottom). Decrease of the estimator $\eta^2 = \eta_\ell^2$, the absolute value of the reliability related consistency error $e_r = |e^c_{\ell,rel}|$ and its heuristic estimate $E_r = |E^{c,2}_{\ell,rel}|$, and the efficiency related consistency error $e_e = e^{c,L^2}_{\ell,eff}$ and its heuristic estimate $E_e = E^c_{\ell,eff}$ as functions of the DOF on a double logarithmic scale for uniform refinement (left) and adaptive refinement (right). For Example 6 the quantities $|e^c_{\ell,rel}|$ and $e^{c,L^2}_{\ell,eff}$ are not given since the exact solution is unknown.

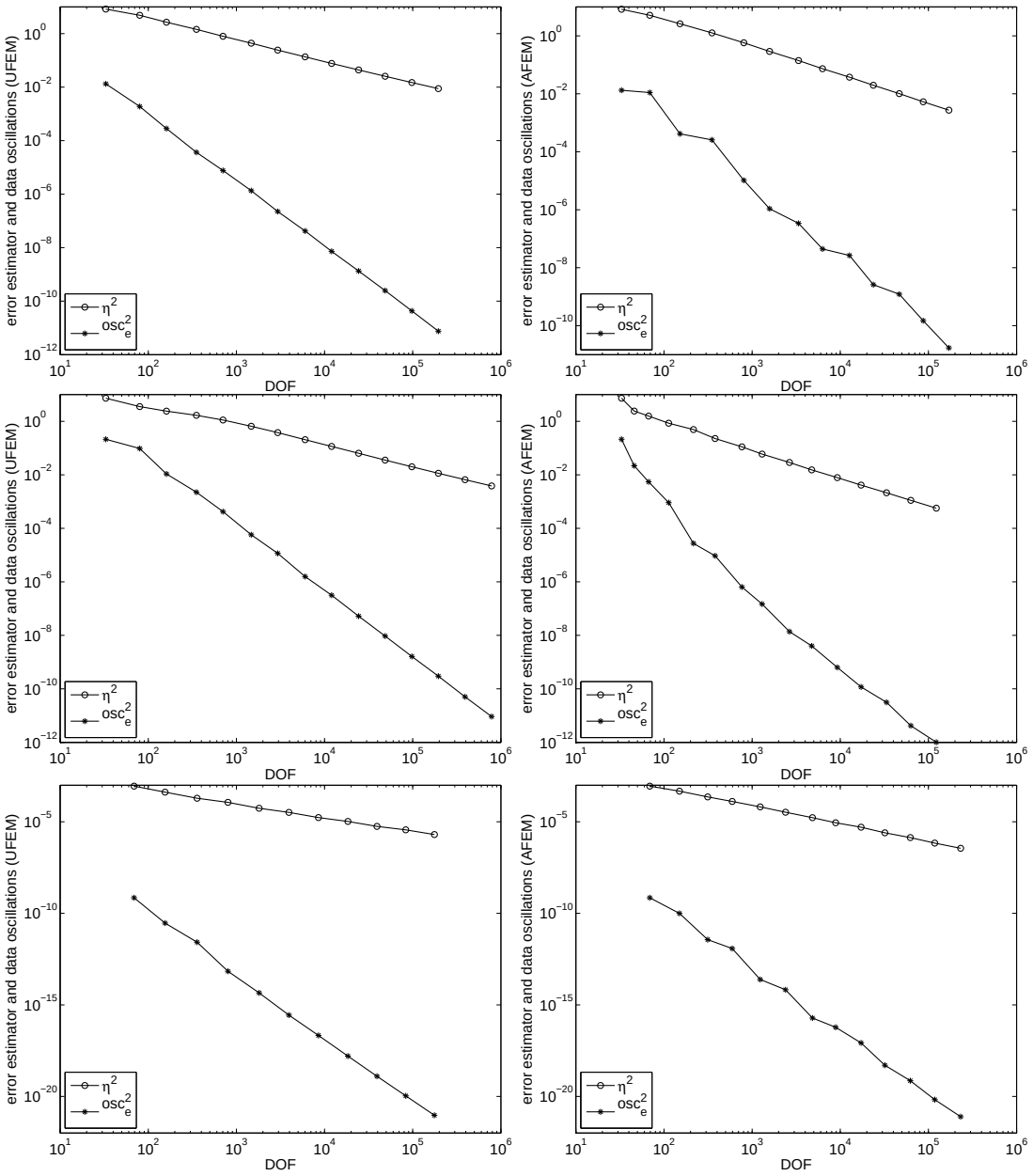


Figure 8.23: Example 1 (top), Example 2 (middle), Example 3 (bottom). Decrease of the estimator $\eta^2 = \eta_\ell^2$ and the efficiency related data oscillation $osc_e^2 = osc_{\ell,eff}^2$ as functions of the DOF on a logarithmic scale for uniform refinement (left) and adaptive refinement (right).

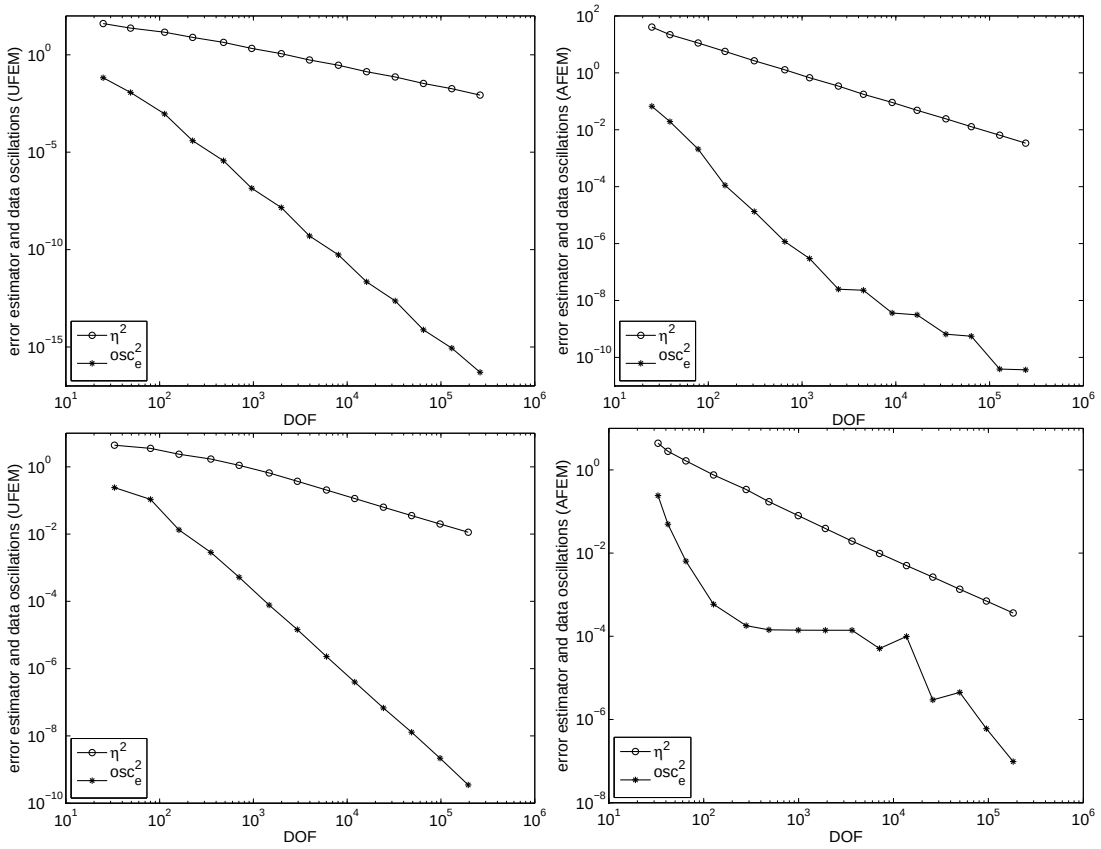


Figure 8.24: Example 4 (top), Example 5 (bottom). Decrease of the estimator $\eta^2 = \eta_\ell^2$ and the efficiency related data oscillation $osc_e^2 = osc_{\ell,eff}^2$ as functions of the DOF on a logarithmic scale for uniform refinement (left) and adaptive refinement (right).

8.4.6 Comparison of adaptive and uniform algorithms in view of genuine computational times

Due to the fact that the PDASS algorithm requires solution of linear systems only for the unknowns in the inactive set $I_\ell^{(k)}$, comparison of the performance of the adaptive and the uniform algorithm in terms of the dependences $\eta_\ell(DOF)$ ($|||e_\ell|||(DOF)$) might lead to wrong conclusions with regard to the benefit of the adaptive refinement. In particular, this can happen if the inactive set is very small.

Examples 1 and 2 are very similar, yet the inactive set $\mathcal{I}^*$ of Example 2 occupies a relatively small subset of Ω. Whereas for Example 1 the computational times t_ℓ corresponding to similar numbers of DOF in case of uniform and adaptive mesh refinement are comparable (see Table A.1.10), for Example 2 they differ dramatically (see Table A.2.10). Note, however, that for similar numbers of *inactive nodal points* N_ℓ^I, the computational times required by the solver on uniform and adaptive meshes are comparable also in case of Example 2. This observation tells us that for the problem under consideration (and, presumably, for contact problems in general), to compare the performance of the uniform and the adaptive mesh refinement, one should analyze error quantities as functions of N_ℓ^I rather than of DOF.

Figure 8.25 shows the decrease of η_ℓ and $|||e_\ell|||$ for Examples 1 and 2 as functions of N_ℓ^I for the uniform and adaptive mesh refinement. Whereas the plot corresponding to Example 1 is almost identical to that we have seen in Paragraph 8.4.2, in case of Example 2 (cf. Figure 8.11) there is an obvious difference. That is, when $|\mathcal{I}^*|$ is comparable to $|\Omega|$, both, dependences on DOF and on N_ℓ^I give approximately the same information, whereas if $|\mathcal{I}^*|/|\Omega| << 1$, comparative analysis of uniform and adaptive refinement should, fairly speaking, rely on error quantities as functions of N_ℓ^I.

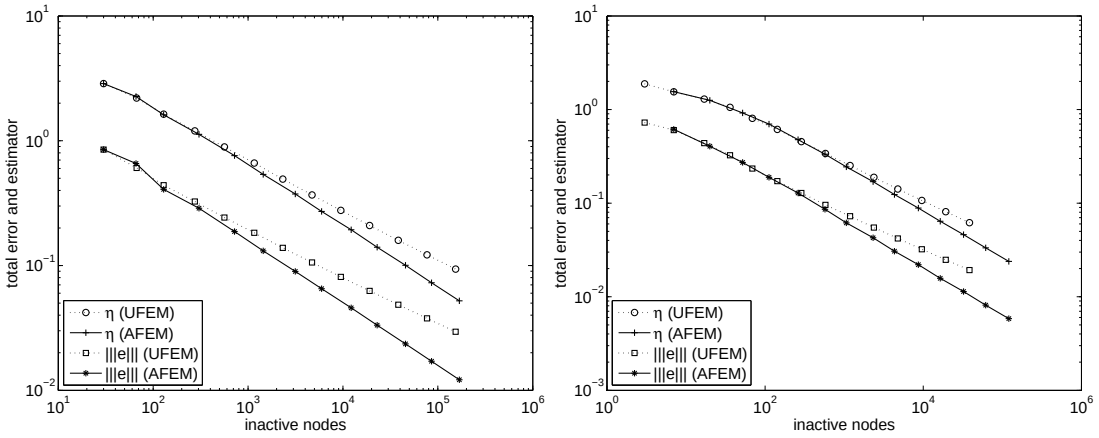


Figure 8.25: Example 1 (left), Example 2 (right). Decrease of the estimator $\eta = \eta_\ell$ and the total error $|||e||| = |||e_\ell|||$ as functions of N_ℓ^I on a double logarithmic scale for uniform (UFEM) and adaptive (AFEM) refinement.

9 Conclusions

The main aim of this work was to develop an adaptive finite element scheme based on the standard residual-type a posteriori error estimator for the optimally controlled elliptic obstacle problem.

In the ideal case, some general convergence result should precede the a posteriori error analysis since convergence to a solution is implicitly assumed by the a posteriori estimates. However, for the problem under consideration with the utilized approximation scheme no convergence result is yet available in the literature to the best of my knowledge. As one of the main results of this thesis, the following convergence result is proved in Chapter 5: for discrete C-stationary points computed on sequences of uniformly refined meshes, under assumption of boundedness of the sequences of the discrete state and control functions, there exists a subsequence converging to an almost C-stationary point of the optimal control problem. One possible suggestion for further research could be an improvement of this convergence result, e.g., as suggested in Remark 5.2.2.

The residual-type a posteriori error estimator η_ℓ suggested in Chapter 6 is shown to be reliable and efficient up to the consistency error terms $e^c_{\ell,rel}$, $e^c_{\ell,eff}$ and data oscillation $osc_{\ell,rel}$, $osc_{\ell,eff}$. In Chapter 7, an adaptive finite element algorithm is suggested that uses the computable part of the a posteriori error estimate, η_ℓ and osc_ℓ, to adapt the mesh to the structure of the unknown solution. Unfortunately, it is impossible to include the consistency errors into the refinement process since $e^c_{\ell,rel}$ and $e^c_{\ell,eff}$ are in general not computable. Therefore, it is assumed that $e^c_{\ell,rel}$, $e^c_{\ell,eff}$ are negligibly small compared to η_ℓ and do not contain any information essential for proper adaptive mesh refinement and the adaptive algorithm simply ignores these terms.

In order to verify this hypothesis, heuristic estimates for the uncomputable terms $e^c_{\ell,rel}$, $e^c_{\ell,eff}$ were suggested in Section 6.6. Some basic questions had to be investigated in order to make the latter estimation possible and, furthermore, to make assumptions that $e^c_{\ell,rel}$ and $e^c_{\ell,eff}$ are small sound plausible. For instance, local properties of V^*-functionals and localization of dual pairings have been considered in detail and the local structure of the stationary points has been analyzed.

In order to test the adaptive algorithm in comparison to the uniform algorithm, extensive numerical simulations were performed. This also allowed to analyze the behavior of the reliability and efficiency related consistency errors and the associated heuristic estimates.

The numerical examples clearly show the benefit of the adaptive algorithm compared to the uniform algorithm which can be seen in enhanced convergence rates. This result justifies the use of the adaptive algorithm in general and of the residual a posteriori error estimates developed in this work in particular. As a positive side effect, the adaptive mesh refinement also seems to suppress chattering of the solver at least at some iterations. It should also be noted that the adaptive mesh refinement does not resolve the strongly active set $\mathcal{C}$ well. This, however, has no noticeable influence on the performance of the adaptive method. The active set is typically well approximated by the adaptive algorithm.

For both, uniform and adaptive mesh refinement, the exact consistency errors are dominated by the error estimator in all examples with known analytical solution, which confirms the assumption that the consistency errors are small.

Concerning the heuristic estimates of consistency errors, the results are diverse. The sharp estimation of the efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ by its heuristic estimate in case of the examples with known analytical solution makes the predictions made by the estimate look credible. From three types of heuristic estimates for the reliability related consistency error $e^{c}_{\ell,rel}$ proposed, it seems that $E^{c,1}_{\ell,rel}$ (if the data is smooth) and $E^{c,2}_{\ell,rel}$ perform well for both, problems with and without strict complementarity. However, for some examples with lack of strict complementarity they are not able to capture the decay rates correctly. This deficiency can presumably be fixed by employing a different solver.

The performance of heuristic approximations of sets $\bar{\mathcal{A}}_\ell$ and $\bar{\mathcal{C}}_\ell$ has also been analyzed. The data show that, on one hand, $\bar{\mathcal{A}}_\ell$ ($\bar{\mathcal{C}}_\ell$) does not provide better approximations to $\mathcal{A}$ ($\mathcal{C}$) than the discrete set $\mathcal{A}_\ell$ ($\mathcal{C}_\ell$) does. This could presumably be fixed by varying the parameters in the definitions of $\bar{\mathcal{A}}_\ell$ and $\bar{\mathcal{C}}_\ell$. On the other hand, the L^1-norm of the difference between the characteristic functions of $\bar{\mathcal{A}}_\ell$ ($\bar{\mathcal{C}}_\ell$) and $\mathcal{A}_\ell$ ($\mathcal{C}_\ell$) typically provides an upper bound with correct decay rates to the L^1-norm of the difference between the characteristic functions of $\mathcal{A}$ ($\mathcal{C}$) and $\mathcal{A}_\ell$ ($\mathcal{C}_\ell$). Overall, the heuristic set approximations provide a satisfactory basis for the heuristic consistency error estimates.

An important unresolved issue is the lack of a theory that in view of the non-uniqueness of the solutions would guarantee convergence of numerical schemes to a certain solution as the mesh is refined. In all examples presented in Chapter 8, the numerical scheme converges to one and the same solution. Moreover, in case of the examples with known exact solutions, convergence to the expected solution is observed. However, it is possible to construct examples where this is not the case. For instance, by using the solver PDASS(LSC,ϵ) for examples with known exact solution where $\mu \neq 0$ in $\mathcal{B}$. In this case the algorithm enforces $\mu_\ell = 0$ in B_ℓ and converges to a solution that differs from the expected one.

A Tabulated results of the numerical experiments

This Appendix supplements the documentation of the numerical results presented in Chapter 8.

A.1 Example 1

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$2.87e+00$	$1.85e+00$	$2.20e+00$
1	$2.20e+00$	$1.44e+00$	$1.66e+00$
2	$1.63e+00$	$1.08e+00$	$1.22e+00$
3	$1.20e+00$	$8.02e-01$	$8.92e-01$
4	$8.93e-01$	$6.01e-01$	$6.60e-01$
5	$6.62e-01$	$4.48e-01$	$4.88e-01$
6	$4.93e-01$	$3.35e-01$	$3.61e-01$
7	$3.68e-01$	$2.51e-01$	$2.69e-01$
8	$2.77e-01$	$1.90e-01$	$2.01e-01$
9	$2.09e-01$	$1.44e-01$	$1.52e-01$
10	$1.59e-01$	$1.10e-01$	$1.15e-01$
11	$1.22e-01$	$8.45e-02$	$8.77e-02$
12	$9.36e-02$	$6.52e-02$	$6.73e-02$

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$2.87e+00$	$1.85e+00$	$2.20e+00$
1	$2.26e+00$	$1.52e+00$	$1.67e+00$
2	$1.62e+00$	$1.08e+00$	$1.21e+00$
3	$1.13e+00$	$7.60e-01$	$8.30e-01$
4	$7.60e-01$	$5.12e-01$	$5.61e-01$
5	$5.38e-01$	$3.63e-01$	$3.97e-01$
6	$3.75e-01$	$2.52e-01$	$2.77e-01$
7	$2.71e-01$	$1.84e-01$	$2.00e-01$
8	$1.93e-01$	$1.31e-01$	$1.43e-01$
9	$1.40e-01$	$9.50e-02$	$1.03e-01$
10	$1.01e-01$	$6.79e-02$	$7.41e-02$
11	$7.28e-02$	$4.93e-02$	$5.36e-02$
12	$5.22e-02$	$3.52e-02$	$3.85e-02$

Table A.1.1: Example 1. Error estimator η_ℓ and its components $\eta_\ell(y)$ and $\eta_\ell(p)$ for uniform refinement (left) and adaptive refinement (right).

ℓ	DOF	$\|\|\|e_\ell\|\|\|$	$\|e_{\ell,u}\|_{0,\Omega}$	$\|e_{\ell,y}\|_{1,\Omega}$	$\|e_{\ell,p}\|_{1,\Omega}$
0	33	$8.50e-01$	$1.08e-01$	$6.01e-01$	$5.92e-01$
1	80	$6.05e-01$	$5.69e-02$	$4.29e-01$	$4.24e-01$
2	161	$4.40e-01$	$3.10e-02$	$3.12e-01$	$3.10e-01$
3	352	$3.26e-01$	$1.74e-02$	$2.31e-01$	$2.30e-01$
4	705	$2.43e-01$	$1.00e-02$	$1.72e-01$	$1.71e-01$
5	1472	$1.83e-01$	$5.92e-03$	$1.29e-01$	$1.29e-01$
6	2945	$1.39e-01$	$3.53e-03$	$9.80e-02$	$9.80e-02$
7	6016	$1.06e-01$	$2.13e-03$	$7.48e-02$	$7.48e-02$
8	12033	$8.12e-02$	$1.30e-03$	$5.74e-02$	$5.74e-02$
9	24320	$6.26e-02$	$7.99e-04$	$4.43e-02$	$4.43e-02$
10	48641	$4.85e-02$	$4.94e-04$	$3.43e-02$	$3.43e-02$
11	97792	$3.78e-02$	$3.07e-04$	$2.67e-02$	$2.67e-02$
12	195585	$2.95e-02$	$1.91e-04$	$2.09e-02$	$2.09e-02$

Table A.1.2: Example 1. Number of DOF, total error $|||e_\ell|||$ and its components $\|e_{\ell,u}\|_{0,\Omega}$, $\|e_{\ell,y}\|_{1,\Omega}$, and $\|e_{\ell,p}\|_{1,\Omega}$ for uniform refinement.

ℓ	DOF	$\|\|\|e_\ell\|\|\|$	$\|e_{\ell,u}\|_{0,\Omega}$	$\|e_{\ell,y}\|_{1,\Omega}$	$\|e_{\ell,p}\|_{1,\Omega}$
0	33	$8.50e-01$	$1.08e-01$	$6.01e-01$	$5.92e-01$
1	69	$6.57e-01$	$6.50e-02$	$4.65e-01$	$4.60e-01$
2	152	$4.08e-01$	$2.40e-02$	$2.88e-01$	$2.87e-01$
3	350	$2.88e-01$	$1.30e-02$	$2.03e-01$	$2.03e-01$
4	805	$1.87e-01$	$5.32e-03$	$1.33e-01$	$1.32e-01$
5	1576	$1.31e-01$	$2.55e-03$	$9.28e-02$	$9.27e-02$
6	3361	$8.98e-02$	$1.21e-03$	$6.35e-02$	$6.35e-02$
7	6289	$6.52e-02$	$6.41e-04$	$4.61e-02$	$4.61e-02$
8	12726	$4.58e-02$	$3.20e-04$	$3.24e-02$	$3.24e-02$
9	23708	$3.33e-02$	$1.68e-04$	$2.35e-02$	$2.35e-02$
10	46719	$2.36e-02$	$8.46e-05$	$1.67e-02$	$1.67e-02$
11	87342	$1.71e-02$	$4.46e-05$	$1.21e-02$	$1.21e-02$
12	170785	$1.21e-02$	$2.24e-05$	$8.59e-03$	$8.59e-03$

Table A.1.3: Example 1. Number of DOF, total error $|||e_\ell|||$ and its components $\|e_{\ell,u}\|_{0,\Omega}$, $\|e_{\ell,y}\|_{1,\Omega}$, and $\|e_{\ell,p}\|_{1,\Omega}$ for adaptive refinement.

ℓ	$osc_{\ell,eff}$	$osc_{\mathcal{Z}_\ell}(f)$	$osc_{\mathcal{I}_\ell}(y^d)$	$osc_{\ell,rel}$
0	$1.15e-01$	$7.73e-02$	$8.52e-02$	$0.00e+00$
1	$4.35e-02$	$3.01e-02$	$3.14e-02$	$0.00e+00$
2	$1.68e-02$	$1.18e-02$	$1.20e-02$	$0.00e+00$
3	$6.04e-03$	$4.25e-03$	$4.30e-03$	$0.00e+00$
4	$2.76e-03$	$1.95e-03$	$1.96e-03$	$0.00e+00$
5	$1.16e-03$	$8.19e-04$	$8.21e-04$	$0.00e+00$
6	$4.73e-04$	$3.34e-04$	$3.35e-04$	$0.00e+00$
7	$2.05e-04$	$1.45e-04$	$1.45e-04$	$0.00e+00$
8	$8.54e-05$	$6.04e-05$	$6.04e-05$	$0.00e+00$
9	$3.65e-05$	$2.58e-05$	$2.58e-05$	$0.00e+00$
10	$1.58e-05$	$1.11e-05$	$1.11e-05$	$0.00e+00$
11	$6.56e-06$	$4.64e-06$	$4.64e-06$	$0.00e+00$
12	$2.75e-06$	$1.94e-06$	$1.94e-06$	$0.00e+00$

Table A.1.4: Example 1. Efficiency related data oscillation $osc_{\ell,eff}$, its components $osc_{\mathcal{Z}_\ell}(f)$, $osc_{\mathcal{I}_\ell}(y^d)$, and reliability related data oscillation $osc_{\ell,rel}$ for uniform refinement.

ℓ	$osc_{\ell,eff}$	$osc_{\mathcal{Z}_\ell}(f)$	$osc_{\mathcal{I}_\ell}(y^d)$	$osc_{\ell,rel}$
0	$1.15e-01$	$7.73e-02$	$8.52e-02$	$0.00e+00$
1	$1.05e-01$	$7.16e-02$	$7.67e-02$	$0.00e+00$
2	$2.04e-02$	$1.43e-02$	$1.45e-02$	$0.00e+00$
3	$1.60e-02$	$1.13e-02$	$1.13e-02$	$0.00e+00$
4	$3.22e-03$	$2.28e-03$	$2.28e-03$	$0.00e+00$
5	$1.04e-03$	$7.36e-04$	$7.36e-04$	$0.00e+00$
6	$5.83e-04$	$4.12e-04$	$4.12e-04$	$0.00e+00$
7	$2.10e-04$	$1.49e-04$	$1.49e-04$	$0.00e+00$
8	$1.63e-04$	$1.15e-04$	$1.15e-04$	$0.00e+00$
9	$5.10e-05$	$3.60e-05$	$3.60e-05$	$0.00e+00$
10	$3.49e-05$	$2.47e-05$	$2.47e-05$	$0.00e+00$
11	$1.22e-05$	$8.61e-06$	$8.61e-06$	$0.00e+00$
12	$4.14e-06$	$2.93e-06$	$2.93e-06$	$0.00e+00$

Table A.1.5: Example 1. Efficiency related data oscillation $osc_{\ell,eff}$, its components $osc_{\mathcal{Z}_\ell}(f)$, $osc_{\mathcal{I}_\ell}(y^d)$, and reliability related data oscillation $osc_{\ell,rel}$ for adaptive refinement.

ℓ	$e^c_{\ell,rel}$	$E^{c,2}_{\ell,rel}$	$E^{c,3}_{\ell,rel}$	$e^{c,L^2}_{\ell,eff}$	$E^c_{\ell,eff}$
0	$7.10e-03$	$3.60e+00$	$-3.96e-02$	$2.75e-01$	$1.63e-01$
1	$-3.36e-03$	$5.15e+00$	$-8.40e-03$	$1.21e-01$	$7.65e-02$
2	$2.17e-03$	$2.05e+00$	$-6.23e-03$	$3.00e-02$	$2.56e-02$
3	$9.59e-04$	$7.60e-01$	$-3.52e-03$	$8.31e-03$	$5.39e-03$
4	$4.17e-04$	$1.82e-01$	$-1.56e-03$	$3.67e-03$	$3.73e-03$
5	$1.52e-04$	$6.89e-02$	$-4.99e-04$	$1.50e-03$	$1.18e-03$
6	$6.51e-05$	$2.16e-02$	$-2.12e-04$	$4.26e-04$	$4.07e-04$
7	$2.58e-05$	$8.05e-03$	$-9.40e-05$	$1.49e-04$	$1.28e-04$
8	$9.17e-06$	$2.53e-03$	$-3.29e-05$	$5.84e-05$	$5.71e-05$
9	$3.55e-06$	$9.63e-04$	$-1.13e-05$	$2.18e-05$	$2.06e-05$
10	$1.12e-06$	$3.09e-04$	$-4.22e-06$	$8.23e-06$	$8.21e-06$
11	$4.77e-07$	$1.14e-04$	$-1.58e-06$	$2.61e-06$	$2.67e-06$
12	$1.51e-07$	$3.67e-05$	$-5.62e-07$	$9.58e-07$	$9.90e-07$

Table A.1.6: Example 1. Reliability related consistency error $e^c_{\ell,rel}$ with heuristic estimates $E^{c,2}_{\ell,rel}$, $E^{c,3}_{\ell,rel}$ and efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ with heuristic estimate $E^c_{\ell,eff}$ for uniform refinement.

ℓ	$e^c_{\ell,rel}$	$E^{c,2}_{\ell,rel}$	$E^{c,3}_{\ell,rel}$	$e^{c,L^2}_{\ell,eff}$	$E^c_{\ell,eff}$
0	$7.10e-03$	$3.60e+00$	$-3.96e-02$	$2.75e-01$	$1.63e-01$
1	$5.91e-03$	$4.28e+00$	$-3.40e-02$	$2.75e-01$	$1.58e-01$
2	$-1.39e-03$	$1.73e+00$	$-8.75e-03$	$3.08e-02$	$1.47e-02$
3	$2.46e-04$	$1.05e+00$	$-4.16e-03$	$1.46e-02$	$8.07e-03$
4	$1.23e-04$	$3.94e-01$	$-1.49e-03$	$3.28e-03$	$3.19e-03$
5	$8.63e-05$	$5.88e-02$	$-3.11e-04$	$1.56e-03$	$1.45e-03$
6	$3.77e-05$	$4.21e-02$	$-1.34e-04$	$4.61e-04$	$2.93e-04$
7	$1.71e-05$	$1.19e-02$	$-5.21e-05$	$2.11e-04$	$1.70e-04$
8	$5.44e-06$	$4.39e-03$	$-2.25e-05$	$9.27e-05$	$8.24e-05$
9	$2.12e-06$	$1.65e-03$	$-7.92e-06$	$2.79e-05$	$2.46e-05$
10	$1.07e-06$	$7.19e-04$	$-3.33e-06$	$1.02e-05$	$1.11e-05$
11	$3.98e-07$	$2.27e-04$	$-1.09e-06$	$4.00e-06$	$3.46e-06$
12	$1.75e-07$	$1.08e-04$	$-4.78e-07$	$1.30e-06$	$1.42e-06$

Table A.1.7: Example 1. Reliability related consistency error $e^c_{\ell,rel}$ with heuristic estimates $E^{c,2}_{\ell,rel}$, $E^{c,3}_{\ell,rel}$ and efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ with heuristic estimate $E^c_{\ell,eff}$ for adaptive refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$	$\|\|\|e_\ell\|\|\|$	$\|\|e_{\ell,u}\|\|_{0,\Omega}$	$\|\|e_{\ell,y}\|\|_{1,\Omega}$	$\|\|e_{\ell,p}\|\|_{1,\Omega}$
1	0.30	0.28	0.32	0.38	0.72	0.38	0.38
2	0.43	0.41	0.44	0.46	0.87	0.46	0.45
3	0.39	0.38	0.40	0.38	0.74	0.38	0.38
4	0.43	0.42	0.43	0.43	0.80	0.43	0.42
5	0.41	0.40	0.41	0.38	0.71	0.38	0.38
6	0.43	0.42	0.43	0.40	0.75	0.40	0.40
7	0.41	0.40	0.41	0.38	0.71	0.38	0.38
8	0.41	0.41	0.42	0.38	0.71	0.38	0.38
9	0.40	0.39	0.40	0.37	0.69	0.37	0.37
10	0.39	0.39	0.40	0.37	0.69	0.37	0.37
11	0.38	0.38	0.39	0.36	0.68	0.36	0.36
12	0.38	0.37	0.38	0.36	0.68	0.36	0.36

Table A.1.8: Example 1. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, $\eta_\ell(p)$, $|||e_\ell|||$, $||e_{\ell,u}||_{0,\Omega}$, $||e_{\ell,y}||_{1,\Omega}$, and $||e_{\ell,p}||_{1,\Omega}$ for uniform refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$	$\|\|\|e_\ell\|\|\|$	$\|\|e_{\ell,u}\|\|_{0,\Omega}$	$\|\|e_{\ell,y}\|\|_{1,\Omega}$	$\|\|e_{\ell,p}\|\|_{1,\Omega}$
1	0.32	0.26	0.37	0.35	0.68	0.35	0.34
2	0.43	0.44	0.41	0.60	1.26	0.60	0.60
3	0.43	0.42	0.45	0.42	0.73	0.42	0.42
4	0.47	0.47	0.47	0.51	1.08	0.51	0.51
5	0.51	0.51	0.52	0.53	1.10	0.53	0.53
6	0.48	0.48	0.47	0.50	0.99	0.50	0.50
7	0.51	0.51	0.52	0.51	1.01	0.51	0.51
8	0.48	0.48	0.48	0.50	0.98	0.50	0.50
9	0.52	0.51	0.52	0.52	1.03	0.52	0.52
10	0.49	0.50	0.49	0.51	1.02	0.51	0.51
11	0.51	0.51	0.52	0.51	1.02	0.51	0.51
12	0.50	0.50	0.49	0.51	1.02	0.51	0.51

Table A.1.9: Example 1. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, $\eta_\ell(p)$, $|||e_\ell|||$, $||e_{\ell,u}||_{0,\Omega}$, $||e_{\ell,y}||_{1,\Omega}$, and $||e_{\ell,p}||_{1,\Omega}$ for adaptive refinement.

ℓ	N_ℓ^I	N_ℓ^A	k'_ℓ	t_ℓ
0	30	3	3	21”
1	67	13	4	42”
2	129	32	4	1’ 20”
3	275	77	5	2’ 35”
4	563	142	6	5’ 12”
5	1162	310	7	10’ 26”
6	2324	621	9	20’ 44”
7	4723	1293	11	41’ 27”
8	9490	2543	16	82’ 60”
9	19106	5214	19	166’ 01”
10	38284	10357	30	332’ 01”
11	76814	20978	34	663’ 42”
12	153778	41807	58	1327’ 18”

ℓ	N_ℓ^I	N_ℓ^A	k'_ℓ	t_ℓ
0	30	3	3	21”
1	66	3	3	1’ 06”
2	129	23	4	1’ 43”
3	304	46	5	4’ 00”
4	720	85	8	9’ 11”
5	1443	133	8	15’ 57”
6	3143	218	12	37’ 02”
7	5955	334	13	65’ 32”
8	12220	506	20	141’ 05”
9	22940	768	21	250’ 27”
10	45599	1120	37	516’ 55”
11	85743	1599	39	940’ 40”
12	168431	2354	68	1877’ 23”

Table A.1.10: Example 1. Number of inactive N_ℓ^I and active nodal points N_ℓ^A, number of iterations of PDASS k'_ℓ and computational time t_ℓ in minutes and seconds for uniform refinement (left) and adaptive refinement (right).

A.2 Example 2

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$2.70e+00$	$1.89e+00$	$1.93e+00$
1	$1.88e+00$	$1.32e+00$	$1.34e+00$
2	$1.55e+00$	$1.08e+00$	$1.11e+00$
3	$1.29e+00$	$9.06e-01$	$9.20e-01$
4	$1.06e+00$	$7.42e-01$	$7.50e-01$
5	$8.06e-01$	$5.67e-01$	$5.72e-01$
6	$6.14e-01$	$4.33e-01$	$4.36e-01$
7	$4.54e-01$	$3.20e-01$	$3.22e-01$
8	$3.39e-01$	$2.39e-01$	$2.41e-01$
9	$2.52e-01$	$1.78e-01$	$1.79e-01$
10	$1.89e-01$	$1.33e-01$	$1.34e-01$
11	$1.41e-01$	$9.99e-02$	$1.00e-01$
12	$1.07e-01$	$7.54e-02$	$7.56e-02$
13	$8.10e-02$	$5.72e-02$	$5.74e-02$
14	$6.18e-02$	$4.37e-02$	$4.38e-02$

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$2.70e+00$	$1.89e+00$	$1.93e+00$
1	$1.55e+00$	$1.09e+00$	$1.11e+00$
2	$1.25e+00$	$8.82e-01$	$8.91e-01$
3	$9.20e-01$	$6.49e-01$	$6.52e-01$
4	$7.00e-01$	$4.93e-01$	$4.96e-01$
5	$4.77e-01$	$3.36e-01$	$3.38e-01$
6	$3.33e-01$	$2.35e-01$	$2.36e-01$
7	$2.44e-01$	$1.72e-01$	$1.73e-01$
8	$1.70e-01$	$1.20e-01$	$1.20e-01$
9	$1.24e-01$	$8.72e-02$	$8.76e-02$
10	$8.86e-02$	$6.25e-02$	$6.28e-02$
11	$6.41e-02$	$4.52e-02$	$4.54e-02$
12	$4.61e-02$	$3.25e-02$	$3.27e-02$
13	$3.33e-02$	$2.35e-02$	$2.36e-02$
14	$2.38e-02$	$1.68e-02$	$1.69e-02$

Table A.2.1: Example 2. Error estimator η_ℓ and its components $\eta_\ell(y)$ and $\eta_\ell(p)$ for uniform refinement (left) and adaptive refinement (right).

ℓ	DOF	$\|\|\|e_\ell\|\|\|$	$\|e_{\ell,u}\|_{0,\Omega}$	$\|e_{\ell,y}\|_{1,\Omega}$	$\|e_{\ell,p}\|_{1,\Omega}$
0	33	$7.78e-01$	$6.93e-02$	$5.48e-01$	$5.48e-01$
1	80	$7.25e-01$	$6.77e-02$	$5.11e-01$	$5.11e-01$
2	161	$6.04e-01$	$4.21e-02$	$4.26e-01$	$4.26e-01$
3	352	$4.38e-01$	$2.31e-02$	$3.09e-01$	$3.09e-01$
4	705	$3.25e-01$	$1.26e-02$	$2.30e-01$	$2.29e-01$
5	1472	$2.34e-01$	$6.83e-03$	$1.65e-01$	$1.65e-01$
6	2945	$1.72e-01$	$3.86e-03$	$1.21e-01$	$1.21e-01$
7	6016	$1.28e-01$	$2.15e-03$	$9.07e-02$	$9.07e-02$
8	12033	$9.58e-02$	$1.26e-03$	$6.77e-02$	$6.77e-02$
9	24320	$7.24e-02$	$7.35e-04$	$5.12e-02$	$5.12e-02$
10	48641	$5.49e-02$	$4.38e-04$	$3.88e-02$	$3.88e-02$
11	97792	$4.19e-02$	$2.64e-04$	$2.97e-02$	$2.97e-02$
12	195585	$3.22e-02$	$1.60e-04$	$2.28e-02$	$2.28e-02$
13	392192	$2.48e-02$	$9.81e-05$	$1.76e-02$	$1.76e-02$
14	784385	$1.93e-02$	$6.05e-05$	$1.36e-02$	$1.36e-02$

Table A.2.2: Example 2. Number of DOF, total error $|||e_\ell|||$ and its components $\|e_{\ell,u}\|_{0,\Omega}$, $\|e_{\ell,y}\|_{1,\Omega}$, and $\|e_{\ell,p}\|_{1,\Omega}$ for uniform refinement.

ℓ	DOF	$\|\|\|e_\ell\|\|\|$	$\|e_{\ell,u}\|_{0,\Omega}$	$\|e_{\ell,y}\|_{1,\Omega}$	$\|e_{\ell,p}\|_{1,\Omega}$
0	33	$7.78e-01$	$6.93e-02$	$5.48e-01$	$5.48e-01$
1	46	$6.15e-01$	$4.18e-02$	$4.34e-01$	$4.34e-01$
2	67	$4.05e-01$	$1.86e-02$	$2.86e-01$	$2.86e-01$
3	113	$2.72e-01$	$8.57e-03$	$1.92e-01$	$1.92e-01$
4	216	$1.89e-01$	$4.32e-03$	$1.33e-01$	$1.33e-01$
5	380	$1.28e-01$	$1.96e-03$	$9.07e-02$	$9.07e-02$
6	767	$8.60e-02$	$8.77e-04$	$6.08e-02$	$6.08e-02$
7	1297	$6.16e-02$	$4.64e-04$	$4.35e-02$	$4.35e-02$
8	2652	$4.27e-02$	$2.13e-04$	$3.02e-02$	$3.02e-02$
9	4746	$3.06e-02$	$1.14e-04$	$2.17e-02$	$2.17e-02$
10	9263	$2.20e-02$	$5.69e-05$	$1.56e-02$	$1.56e-02$
11	17237	$1.57e-02$	$3.02e-05$	$1.11e-02$	$1.11e-02$
12	33468	$1.14e-02$	$1.52e-05$	$8.04e-03$	$8.04e-03$
13	63066	$8.13e-03$	$8.04e-06$	$5.75e-03$	$5.75e-03$
14	122894	$5.84e-03$	$4.03e-06$	$4.13e-03$	$4.13e-03$

Table A.2.3: Example 2. Number of DOF, total error $|||e_\ell|||$ and its components $\|e_{\ell,u}\|_{0,\Omega}$, $\|e_{\ell,y}\|_{1,\Omega}$, and $\|e_{\ell,p}\|_{1,\Omega}$ for adaptive refinement.

ℓ	$osc_{\ell,eff}$	$osc_{\mathcal{Z}_\ell}(f)$	$osc_{\mathcal{I}_\ell}(y^d)$	$osc_{\ell,rel}$
0	$4.61e-01$	$3.23e-01$	$3.29e-01$	$0.00e+00$
1	$3.10e-01$	$2.18e-01$	$2.21e-01$	$0.00e+00$
2	$1.04e-01$	$7.29e-02$	$7.38e-02$	$0.00e+00$
3	$4.73e-02$	$3.33e-02$	$3.36e-02$	$0.00e+00$
4	$2.05e-02$	$1.45e-02$	$1.45e-02$	$0.00e+00$
5	$7.60e-03$	$5.37e-03$	$5.37e-03$	$0.00e+00$
6	$3.41e-03$	$2.41e-03$	$2.41e-03$	$0.00e+00$
7	$1.25e-03$	$8.87e-04$	$8.87e-04$	$0.00e+00$
8	$5.61e-04$	$3.96e-04$	$3.96e-04$	$0.00e+00$
9	$2.28e-04$	$1.61e-04$	$1.61e-04$	$0.00e+00$
10	$9.70e-05$	$6.86e-05$	$6.86e-05$	$0.00e+00$
11	$4.02e-05$	$2.84e-05$	$2.84e-05$	$0.00e+00$
12	$1.73e-05$	$1.22e-05$	$1.22e-05$	$0.00e+00$
13	$7.11e-06$	$5.03e-06$	$5.03e-06$	$0.00e+00$
14	$3.03e-06$	$2.14e-06$	$2.14e-06$	$0.00e+00$

Table A.2.4: Example 2. Efficiency related data oscillation $osc_{\ell,eff}$, its components $osc_{\mathcal{Z}_\ell}(f)$, $osc_{\mathcal{I}_\ell}(y^d)$, and reliability related data oscillation $osc_{\ell,rel}$ for uniform refinement.

ℓ	$osc_{\ell,eff}$	$osc_{\mathcal{Z}_\ell}(f)$	$osc_{\mathcal{I}_\ell}(y^d)$	$osc_{\ell,rel}$
0	$4.61e-01$	$3.23e-01$	$3.29e-01$	$0.00e+00$
1	$1.47e-01$	$1.03e-01$	$1.05e-01$	$0.00e+00$
2	$7.39e-02$	$5.21e-02$	$5.24e-02$	$0.00e+00$
3	$3.02e-02$	$2.13e-02$	$2.14e-02$	$0.00e+00$
4	$5.24e-03$	$3.70e-03$	$3.71e-03$	$0.00e+00$
5	$3.06e-03$	$2.16e-03$	$2.16e-03$	$0.00e+00$
6	$7.99e-04$	$5.65e-04$	$5.65e-04$	$0.00e+00$
7	$3.84e-04$	$2.72e-04$	$2.72e-04$	$0.00e+00$
8	$1.17e-04$	$8.30e-05$	$8.30e-05$	$0.00e+00$
9	$6.33e-05$	$4.48e-05$	$4.48e-05$	$0.00e+00$
10	$2.52e-05$	$1.78e-05$	$1.78e-05$	$0.00e+00$
11	$1.09e-05$	$7.68e-06$	$7.68e-06$	$0.00e+00$
12	$5.60e-06$	$3.96e-06$	$3.96e-06$	$0.00e+00$
13	$2.05e-06$	$1.45e-06$	$1.45e-06$	$0.00e+00$
14	$1.00e-06$	$7.10e-07$	$7.10e-07$	$0.00e+00$

Table A.2.5: Example 2. Efficiency related data oscillation $osc_{\ell,eff}$, its components $osc_{\mathcal{Z}_\ell}(f)$, $osc_{\mathcal{I}_\ell}(y^d)$, and reliability related data oscillation $osc_{\ell,rel}$ for adaptive refinement.

ℓ	$e^c_{\ell,rel}$	$E^{c,2}_{\ell,rel}$	$E^{c,3}_{\ell,rel}$	$e^{c,L^2}_{\ell,eff}$	$E^c_{\ell,eff}$
0	$0.00e+00$	$0.00e+00$	$0.00e+00$	$0.00e+00$	$0.00e+00$
1	$4.29e-02$	$1.17e-02$	$-5.17e-03$	$2.03e-02$	$5.03e-02$
2	$-1.30e-03$	$1.94e-01$	$-8.73e-03$	$1.74e-02$	$9.74e-03$
3	$9.71e-03$	$5.06e-02$	$-5.41e-03$	$5.08e-03$	$5.64e-03$
4	$2.07e-03$	$1.31e-02$	$-6.05e-04$	$2.54e-03$	$3.35e-03$
5	$1.35e-03$	$9.12e-03$	$-3.30e-04$	$5.23e-04$	$1.16e-03$
6	$3.43e-04$	$3.17e-03$	$5.34e-05$	$2.92e-04$	$5.85e-04$
7	$2.14e-04$	$1.29e-03$	$2.32e-05$	$6.38e-05$	$1.44e-04$
8	$5.54e-05$	$5.28e-04$	$2.42e-05$	$3.03e-05$	$9.37e-05$
9	$2.56e-05$	$2.17e-04$	$3.92e-06$	$1.07e-05$	$1.84e-05$
10	$5.64e-06$	$6.14e-05$	$2.36e-06$	$4.15e-06$	$6.90e-06$
11	$3.30e-06$	$2.40e-05$	$1.02e-06$	$1.21e-06$	$1.89e-06$
12	$6.10e-07$	$9.07e-06$	$4.03e-07$	$5.36e-07$	$1.12e-06$
13	$3.89e-07$	$3.30e-06$	$1.14e-07$	$1.80e-07$	$2.80e-07$
14	$9.52e-08$	$1.05e-06$	$5.21e-08$	$6.51e-08$	$1.07e-07$

Table A.2.6: Example 2. Reliability related consistency error $e^c_{\ell,rel}$ with heuristic estimates $E^{c,2}_{\ell,rel}$, $E^{c,3}_{\ell,rel}$ and efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ with heuristic estimate $E^c_{\ell,eff}$ for uniform refinement.

ℓ	$e^c_{\ell,rel}$	$E^{c,2}_{\ell,rel}$	$E^{c,3}_{\ell,rel}$	$e^{c,L^2}_{\ell,eff}$	$E^c_{\ell,eff}$
0	$0.00e+00$	$0.00e+00$	$0.00e+00$	$0.00e+00$	$0.00e+00$
1	$3.67e-05$	$1.95e-01$	$-1.09e-02$	$2.08e-02$	$1.49e-02$
2	$5.03e-04$	$3.54e-02$	$-2.69e-03$	$1.24e-02$	$8.02e-03$
3	$2.83e-03$	$5.38e-02$	$-2.36e-03$	$2.08e-03$	$3.00e-03$
4	$8.44e-04$	$4.75e-03$	$-2.52e-04$	$4.67e-04$	$5.70e-04$
5	$1.09e-03$	$4.78e-03$	$-1.40e-04$	$3.25e-04$	$7.67e-04$
6	$6.25e-05$	$6.48e-04$	$-1.29e-06$	$1.10e-04$	$1.05e-04$
7	$1.80e-05$	$4.69e-04$	$-7.75e-06$	$3.64e-05$	$4.53e-05$
8	$4.65e-06$	$8.90e-05$	$-3.68e-08$	$7.80e-06$	$1.17e-05$
9	$2.43e-06$	$5.39e-05$	$-4.10e-07$	$4.22e-06$	$3.92e-06$
10	$9.40e-07$	$2.66e-05$	$-1.52e-07$	$1.66e-06$	$1.79e-06$
11	$3.93e-07$	$1.16e-05$	$-1.28e-07$	$5.72e-07$	$6.40e-07$
12	$1.22e-07$	$5.70e-06$	$-3.81e-08$	$2.13e-07$	$2.91e-07$
13	$5.77e-08$	$2.19e-06$	$-1.36e-08$	$8.30e-08$	$1.04e-07$
14	$1.80e-08$	$6.21e-07$	$-2.98e-09$	$2.68e-08$	$3.45e-08$

Table A.2.7: Example 2. Reliability related consistency error $e^c_{\ell,rel}$ with heuristic estimates $E^{c,2}_{\ell,rel}$, $E^{c,3}_{\ell,rel}$ and efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ with heuristic estimate $E^c_{\ell,eff}$ for adaptive refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$	$\|\|\|e_\ell\|\|\|$	$\|\|e_{\ell,u}\|\|_{0,\Omega}$	$\|\|e_{\ell,y}\|\|_{1,\Omega}$	$\|\|e_{\ell,p}\|\|_{1,\Omega}$
1	0.41	0.41	0.41	0.08	0.03	0.08	0.08
2	0.28	0.28	0.28	0.26	0.68	0.26	0.26
3	0.23	0.23	0.23	0.41	0.76	0.41	0.41
4	0.29	0.29	0.29	0.43	0.88	0.43	0.43
5	0.37	0.36	0.37	0.44	0.83	0.44	0.44
6	0.39	0.39	0.39	0.45	0.82	0.45	0.45
7	0.42	0.42	0.42	0.41	0.82	0.41	0.41
8	0.42	0.42	0.42	0.42	0.78	0.42	0.42
9	0.42	0.42	0.42	0.40	0.76	0.40	0.40
10	0.42	0.42	0.42	0.40	0.75	0.40	0.40
11	0.41	0.41	0.41	0.39	0.73	0.39	0.39
12	0.41	0.41	0.41	0.38	0.72	0.38	0.38
13	0.40	0.40	0.40	0.37	0.71	0.37	0.37
14	0.39	0.39	0.39	0.37	0.70	0.37	0.37

Table A.2.8: Example 2. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, $\eta_\ell(p)$, $|||e_\ell|||$, $||e_{\ell,u}||_{0,\Omega}$, $||e_{\ell,y}||_{1,\Omega}$, and $||e_{\ell,p}||_{1,\Omega}$ for uniform refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$	$\|\|\|e_\ell\|\|\|$	$\|\|e_{\ell,u}\|\|_{0,\Omega}$	$\|\|e_{\ell,y}\|\|_{1,\Omega}$	$\|\|e_{\ell,p}\|\|_{1,\Omega}$
1	1.67	1.67	1.66	0.71	1.53	0.70	0.70
2	0.57	0.55	0.58	1.11	2.15	1.11	1.11
3	0.59	0.59	0.60	0.76	1.48	0.76	0.76
4	0.42	0.42	0.42	0.56	1.06	0.56	0.56
5	0.68	0.68	0.68	0.69	1.40	0.68	0.68
6	0.51	0.51	0.51	0.57	1.15	0.57	0.57
7	0.59	0.59	0.59	0.64	1.21	0.64	0.64
8	0.51	0.51	0.51	0.51	1.09	0.51	0.51
9	0.55	0.55	0.55	0.57	1.07	0.57	0.57
10	0.50	0.50	0.50	0.49	1.04	0.49	0.49
11	0.52	0.52	0.52	0.54	1.02	0.54	0.54
12	0.50	0.50	0.50	0.49	1.03	0.49	0.49
13	0.51	0.51	0.51	0.53	1.01	0.53	0.53
14	0.50	0.50	0.50	0.50	1.03	0.50	0.50

Table A.2.9: Example 2. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, $\eta_\ell(p)$, $|||e_\ell|||$, $||e_{\ell,u}||_{0,\Omega}$, $||e_{\ell,y}||_{1,\Omega}$, and $||e_{\ell,p}||_{1,\Omega}$ for adaptive refinement.

ℓ	N_ℓ^I	N_ℓ^A	k'_ℓ	t_ℓ
0	0	33	2	< 1”
1	3	77	3	01”
2	7	154	4	06”
3	17	335	5	09”
4	36	669	6	18”
5	69	1403	6	34”
6	143	2802	7	1’ 11”
7	287	5729	8	2’ 12”
8	579	11454	9	4’ 27”
9	1187	23133	11	8’ 54”
10	2378	46263	14	17’ 41”
11	4773	93019	18	35’ 07”
12	9581	186004	24	70’ 16”
13	19231	372961	33	140’ 14”
14	38462	745923	46	279’ 52”

ℓ	N_ℓ^I	N_ℓ^A	k'_ℓ	t_ℓ
0	0	33	2	< 1”
1	7	39	4	06”
2	20	47	4	15”
3	52	61	5	45”
4	112	104	7	1’ 11”
5	261	119	9	2’ 57”
6	572	195	9	4’ 51”
7	1064	233	12	11’ 07”
8	2320	332	14	19’ 45”
9	4321	425	21	42’ 59”
10	8667	596	22	71’ 57”
11	16394	843	35	162’ 31”
12	32330	1138	38	267’ 57”
13	61542	1524	62	618’ 53”
14	120703	2191	70	1008’ 15”

Table A.2.10: Example 2. Number of inactive N_ℓ^I and active nodal points N_ℓ^A, number of iterations of PDASS k'_ℓ and computational time t_ℓ in minutes and seconds for uniform refinement (left) and adaptive refinement (right).

A.3 Example 3

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$2.98e-02$	$2.67e-02$	$1.31e-02$
1	$2.04e-02$	$1.85e-02$	$8.68e-03$
2	$1.39e-02$	$1.27e-02$	$5.80e-03$
3	$1.07e-02$	$9.62e-03$	$4.77e-03$
4	$7.39e-03$	$6.71e-03$	$3.11e-03$
5	$5.72e-03$	$5.14e-03$	$2.51e-03$
6	$4.13e-03$	$3.69e-03$	$1.85e-03$
7	$3.24e-03$	$2.92e-03$	$1.41e-03$
8	$2.37e-03$	$2.12e-03$	$1.05e-03$
9	$1.90e-03$	$1.72e-03$	$8.21e-04$
10	$1.42e-03$	$1.26e-03$	$6.57e-04$

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$2.98e-02$	$2.67e-02$	$1.31e-02$
1	$2.17e-02$	$1.94e-02$	$9.84e-03$
2	$1.51e-02$	$1.37e-02$	$6.37e-03$
3	$1.14e-02$	$1.01e-02$	$5.16e-03$
4	$8.08e-03$	$7.27e-03$	$3.51e-03$
5	$5.77e-03$	$5.13e-03$	$2.64e-03$
6	$4.10e-03$	$3.66e-03$	$1.84e-03$
7	$2.98e-03$	$2.65e-03$	$1.37e-03$
8	$2.25e-03$	$1.91e-03$	$1.19e-03$
9	$1.58e-03$	$1.39e-03$	$7.51e-04$
10	$1.17e-03$	$1.01e-03$	$6.06e-04$
11	$8.30e-04$	$7.26e-04$	$4.03e-04$
12	$5.99e-04$	$5.22e-04$	$2.93e-04$

Table A.3.1: Example 3. Error estimator η_ℓ and its components $\eta_\ell(y)$ and $\eta_\ell(p)$ for uniform refinement (left) and adaptive refinement (right).

ℓ	$osc_{\ell,eff}$	$osc_{\mathcal{Z}_\ell}(f)$	$osc_{\mathcal{I}_\ell}(y^d)$	$osc_{\ell,rel}$
0	$2.66e-05$	$1.88e-05$	$1.88e-05$	$0.00e+00$
1	$5.40e-06$	$3.82e-06$	$3.82e-06$	$0.00e+00$
2	$1.63e-06$	$1.15e-06$	$1.15e-06$	$0.00e+00$
3	$2.63e-07$	$1.86e-07$	$1.86e-07$	$0.00e+00$
4	$6.73e-08$	$4.76e-08$	$4.76e-08$	$0.00e+00$
5	$1.66e-08$	$1.18e-08$	$1.18e-08$	$0.00e+00$
6	$4.62e-09$	$3.27e-09$	$3.27e-09$	$0.00e+00$
7	$1.26e-09$	$8.93e-10$	$8.93e-10$	$0.00e+00$
8	$3.56e-10$	$2.51e-10$	$2.51e-10$	$0.00e+00$
9	$1.04e-10$	$7.33e-11$	$7.33e-11$	$0.00e+00$
10	$3.04e-11$	$2.15e-11$	$2.15e-11$	$0.00e+00$

Table A.3.2: Example 3. Efficiency related data oscillation $osc_{\ell,eff}$, its components $osc_{\mathcal{Z}_\ell}(f)$, $osc_{\mathcal{I}_\ell}(y^d)$, and reliability related data oscillation $osc_{\ell,rel}$ for uniform refinement.

ℓ	$osc_{\ell,eff}$	$osc_{\mathcal{Z}_\ell}(f)$	$osc_{\mathcal{I}_\ell}(y^d)$	$osc_{\ell,rel}$
0	$2.66e-05$	$1.88e-05$	$1.88e-05$	$0.00e+00$
1	$9.96e-06$	$7.04e-06$	$7.04e-06$	$0.00e+00$
2	$1.92e-06$	$1.36e-06$	$1.36e-06$	$0.00e+00$
3	$1.10e-06$	$7.75e-07$	$7.75e-07$	$0.00e+00$
4	$1.56e-07$	$1.10e-07$	$1.10e-07$	$0.00e+00$
5	$8.17e-08$	$5.78e-08$	$5.78e-08$	$0.00e+00$
6	$1.38e-08$	$9.78e-09$	$9.78e-09$	$0.00e+00$
7	$7.72e-09$	$5.46e-09$	$5.46e-09$	$0.00e+00$
8	$2.88e-09$	$2.04e-09$	$2.04e-09$	$0.00e+00$
9	$7.09e-10$	$5.01e-10$	$5.01e-10$	$0.00e+00$
10	$2.68e-10$	$1.89e-10$	$1.89e-10$	$0.00e+00$
11	$8.16e-11$	$5.77e-11$	$5.77e-11$	$0.00e+00$
12	$2.79e-11$	$1.97e-11$	$1.97e-11$	$0.00e+00$

Table A.3.3: Example 3. Efficiency related data oscillation $osc_{\ell,eff}$, its components $osc_{\mathcal{Z}_\ell}(f)$, $osc_{\mathcal{I}_\ell}(y^d)$, and reliability related data oscillation $osc_{\ell,rel}$ for adaptive refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
1	0.47	0.46	0.51
2	0.46	0.45	0.48
3	0.32	0.34	0.24
4	0.46	0.44	0.52
5	0.33	0.34	0.28
6	0.42	0.43	0.40
7	0.32	0.31	0.36
8	0.41	0.42	0.38
9	0.29	0.28	0.33
10	0.39	0.41	0.30

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
1	0.41	0.41	0.37
2	0.50	0.47	0.59
3	0.44	0.47	0.33
4	0.47	0.45	0.52
5	0.51	0.53	0.43
6	0.49	0.48	0.51
7	0.52	0.53	0.48
8	0.43	0.49	0.21
9	0.57	0.51	0.74
10	0.45	0.49	0.32
11	0.54	0.51	0.64
12	0.48	0.49	0.47

Table A.3.4: Example 3. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, and $\eta_\ell(p)$ for uniform refinement (left) and adaptive refinement (right).

ℓ	N_ℓ^I	N_ℓ^A	k'_ℓ	t_ℓ
0	54	15	152	6’ 25”
1	121	34	153	8’ 20”
2	283	75	155	22’ 34”
3	669	133	11	47’ 18”
4	1527	284	152	97’ 27”
5	3363	582	202	228’ 34”
6	7305	1224	153	484’ 39”
7	15870	2534	202	1015’ 52”
8	34065	5250	153	2118’ 12”
9	72632	10797	152	4391’ 57”
10	154127	22165	202	9071’ 11”

ℓ	N_ℓ^I	N_ℓ^A	k'_ℓ	t_ℓ
0	54	15	152	6’ 25”
1	128	22	6	9’ 49”
2	270	43	152	27’ 43”
3	558	36	12	54’ 56”
4	1134	101	14	99’ 32”
5	2237	157	153	212’ 11”
6	4515	318	152	400’ 28”
7	8447	422	153	784’ 28”
8	16690	549	202	1430’ 02”
9	30932	1151	153	2791’ 36”
10	60632	1664	152	5063’ 19”
11	114905	3136	152	10167’ 20”
12	227074	4985	153	19013’ 24”

Table A.3.5: Example 3. Number of inactive N_ℓ^I and active nodal points N_ℓ^A, number of iterations of PDASS k'_ℓ and computational time t_ℓ in minutes and seconds for uniform refinement (left) and adaptive refinement (right).

A.4 Example 4

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$6.37e+00$	$4.48e+00$	$4.52e+00$
1	$4.85e+00$	$3.42e+00$	$3.44e+00$
2	$3.81e+00$	$2.69e+00$	$2.70e+00$
3	$2.81e+00$	$1.98e+00$	$1.99e+00$
4	$2.09e+00$	$1.48e+00$	$1.48e+00$
5	$1.46e+00$	$1.03e+00$	$1.03e+00$
6	$1.07e+00$	$7.58e-01$	$7.60e-01$
7	$7.37e-01$	$5.20e-01$	$5.22e-01$
8	$5.40e-01$	$3.81e-01$	$3.83e-01$
9	$3.69e-01$	$2.61e-01$	$2.62e-01$
10	$2.71e-01$	$1.91e-01$	$1.92e-01$
11	$1.85e-01$	$1.30e-01$	$1.31e-01$
12	$1.35e-01$	$9.56e-02$	$9.59e-02$
13	$9.24e-02$	$6.52e-02$	$6.54e-02$

Table A.4.1: Example 4. Error estimator η_ℓ and its components $\eta_\ell(y)$ and $\eta_\ell(p)$ for uniform refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$6.37e+00$	$4.48e+00$	$4.52e+00$
1	$4.69e+00$	$3.31e+00$	$3.32e+00$
2	$3.34e+00$	$2.36e+00$	$2.37e+00$
3	$2.38e+00$	$1.68e+00$	$1.69e+00$
4	$1.63e+00$	$1.15e+00$	$1.15e+00$
5	$1.13e+00$	$7.98e-01$	$8.01e-01$
6	$8.18e-01$	$5.78e-01$	$5.79e-01$
7	$5.84e-01$	$4.13e-01$	$4.14e-01$
8	$4.20e-01$	$2.97e-01$	$2.97e-01$
9	$3.01e-01$	$2.13e-01$	$2.13e-01$
10	$2.19e-01$	$1.55e-01$	$1.55e-01$
11	$1.55e-01$	$1.10e-01$	$1.10e-01$
12	$1.13e-01$	$7.99e-02$	$8.00e-02$
13	$8.02e-02$	$5.67e-02$	$5.68e-02$
14	$5.82e-02$	$4.11e-02$	$4.12e-02$

Table A.4.2: Example 4. Error estimator η_ℓ and its components $\eta_\ell(y)$ and $\eta_\ell(p)$ for adaptive refinement.

ℓ	DOF	$\|\|\|e_\ell\|\|\|$	$\|e_{\ell,u}\|_{0,\Omega}$	$\|e_{\ell,y}\|_{1,\Omega}$	$\|e_{\ell,p}\|_{1,\Omega}$
0	25	$1.65e+00$	$6.39e-02$	$1.17e+00$	$1.16e+00$
1	49	$1.40e+00$	$4.12e-02$	$9.92e-01$	$9.92e-01$
2	113	$1.05e+00$	$2.33e-02$	$7.46e-01$	$7.46e-01$
3	225	$6.92e-01$	$1.02e-02$	$4.90e-01$	$4.90e-01$
4	481	$5.16e-01$	$5.65e-03$	$3.65e-01$	$3.65e-01$
5	961	$3.48e-01$	$2.57e-03$	$2.46e-01$	$2.46e-01$
6	1985	$2.59e-01$	$1.42e-03$	$1.83e-01$	$1.83e-01$
7	3969	$1.75e-01$	$6.42e-04$	$1.23e-01$	$1.23e-01$
8	8065	$1.30e-01$	$3.57e-04$	$9.19e-02$	$9.19e-02$
9	16129	$8.73e-02$	$1.61e-04$	$6.18e-02$	$6.18e-02$
10	32513	$6.50e-02$	$8.92e-05$	$4.60e-02$	$4.60e-02$
11	65025	$4.37e-02$	$4.02e-05$	$3.09e-02$	$3.09e-02$
12	130561	$3.25e-02$	$2.23e-05$	$2.30e-02$	$2.30e-02$
13	261121	$2.18e-02$	$1.00e-05$	$1.54e-02$	$1.54e-02$

Table A.4.3: Example 4. Number of DOF, total error $|||e_\ell|||$ and its components $\|e_{\ell,u}\|_{0,\Omega}$, $\|e_{\ell,y}\|_{1,\Omega}$, and $\|e_{\ell,p}\|_{1,\Omega}$ for uniform refinement.

ℓ	DOF	$\|\|\|e_\ell\|\|\|$	$\|e_{\ell,u}\|_{0,\Omega}$	$\|e_{\ell,y}\|_{1,\Omega}$	$\|e_{\ell,p}\|_{1,\Omega}$
0	25	$1.65e+00$	$6.39e-02$	$1.17e+00$	$1.16e+00$
1	39	$1.30e+00$	$4.90e-02$	$9.17e-01$	$9.15e-01$
2	78	$8.49e-01$	$1.83e-02$	$6.01e-01$	$6.00e-01$
3	151	$5.61e-01$	$7.61e-03$	$3.97e-01$	$3.97e-01$
4	309	$4.02e-01$	$3.84e-03$	$2.84e-01$	$2.84e-01$
5	655	$2.60e-01$	$1.61e-03$	$1.84e-01$	$1.84e-01$
6	1203	$1.96e-01$	$9.44e-04$	$1.38e-01$	$1.38e-01$
7	2439	$1.34e-01$	$4.27e-04$	$9.48e-02$	$9.48e-02$
8	4510	$9.92e-02$	$2.46e-04$	$7.02e-02$	$7.02e-02$
9	9129	$6.90e-02$	$1.15e-04$	$4.88e-02$	$4.88e-02$
10	16851	$5.17e-02$	$6.84e-05$	$3.66e-02$	$3.66e-02$
11	34159	$3.56e-02$	$3.09e-05$	$2.52e-02$	$2.52e-02$
12	63742	$2.65e-02$	$1.78e-05$	$1.87e-02$	$1.87e-02$
13	128066	$1.85e-02$	$8.39e-06$	$1.31e-02$	$1.31e-02$
14	242478	$1.36e-02$	$4.78e-06$	$9.60e-03$	$9.60e-03$

Table A.4.4: Example 4. Number of DOF, total error $|||e_\ell|||$ and its components $\|e_{\ell,u}\|_{0,\Omega}$, $\|e_{\ell,y}\|_{1,\Omega}$, and $\|e_{\ell,p}\|_{1,\Omega}$ for adaptive refinement.

ℓ	$osc_{\ell,eff}$	$osc_{\mathcal{Z}_\ell}(f)$	$osc_{\mathcal{I}_\ell}(y^d)$	$osc_{\ell,rel}$
0	$2.58e-01$	$1.82e-01$	$1.83e-01$	$0.00e+00$
1	$1.08e-01$	$7.65e-02$	$7.68e-02$	$0.00e+00$
2	$3.05e-02$	$2.18e-02$	$2.13e-02$	$0.00e+00$
3	$6.26e-03$	$4.42e-03$	$4.43e-03$	$0.00e+00$
4	$1.91e-03$	$1.36e-03$	$1.33e-03$	$0.00e+00$
5	$3.76e-04$	$2.68e-04$	$2.64e-04$	$0.00e+00$
6	$1.20e-04$	$8.55e-05$	$8.35e-05$	$0.00e+00$
7	$2.23e-05$	$1.62e-05$	$1.54e-05$	$0.00e+00$
8	$7.31e-06$	$5.25e-06$	$5.09e-06$	$0.00e+00$
9	$1.48e-06$	$1.14e-06$	$9.47e-07$	$0.00e+00$
10	$4.80e-07$	$3.65e-07$	$3.12e-07$	$0.00e+00$
11	$8.73e-08$	$6.73e-08$	$5.55e-08$	$0.00e+00$
12	$2.96e-08$	$2.30e-08$	$1.87e-08$	$0.00e+00$
13	$7.07e-09$	$6.16e-09$	$3.47e-09$	$0.00e+00$

Table A.4.5: Example 4. Efficiency related data oscillation $osc_{\ell,eff}$, its components $osc_{\mathcal{Z}_\ell}(f)$, $osc_{\mathcal{I}_\ell}(y^d)$, and reliability related data oscillation $osc_{\ell,rel}$ for uniform refinement.

ℓ	$osc_{\ell,eff}$	$osc_{\mathcal{Z}_\ell}(f)$	$osc_{\mathcal{I}_\ell}(y^d)$	$osc_{\ell,rel}$
0	$2.58e-01$	$1.82e-01$	$1.83e-01$	$0.00e+00$
1	$1.39e-01$	$9.81e-02$	$9.82e-02$	$0.00e+00$
2	$4.57e-02$	$3.23e-02$	$3.23e-02$	$0.00e+00$
3	$1.06e-02$	$7.59e-03$	$7.38e-03$	$0.00e+00$
4	$3.65e-03$	$2.76e-03$	$2.40e-03$	$0.00e+00$
5	$1.09e-03$	$8.36e-04$	$6.93e-04$	$0.00e+00$
6	$5.47e-04$	$4.28e-04$	$3.40e-04$	$0.00e+00$
7	$1.58e-04$	$1.40e-04$	$7.31e-05$	$0.00e+00$
8	$1.52e-04$	$1.32e-04$	$7.50e-05$	$0.00e+00$
9	$6.04e-05$	$5.84e-05$	$1.53e-05$	$0.00e+00$
10	$5.61e-05$	$5.57e-05$	$6.45e-06$	$0.00e+00$
11	$2.56e-05$	$2.55e-05$	$1.84e-06$	$0.00e+00$
12	$2.36e-05$	$2.36e-05$	$1.08e-06$	$0.00e+00$
13	$6.26e-06$	$6.26e-06$	$2.43e-07$	$0.00e+00$
14	$6.06e-06$	$6.06e-06$	$1.34e-07$	$0.00e+00$

Table A.4.6: Example 4. Efficiency related data oscillation $osc_{\ell,eff}$, its components $osc_{\mathcal{Z}_\ell}(f)$, $osc_{\mathcal{I}_\ell}(y^d)$, and reliability related data oscillation $osc_{\ell,rel}$ for adaptive refinement.

ℓ	$e^c_{\ell,rel}$	$E^{c,1}_{\ell,rel}$	$E^{c,2}_{\ell,rel}$	$E^{c,3}_{\ell,rel}$	$e^{c,L^2}_{\ell,eff}$	$E^c_{\ell,eff}$
0	$-2.47e-03$	$1.02e-01$	$1.45e-01$	$-1.21e-03$	$7.82e-04$	$6.35e-04$
1	$8.31e-03$	$7.24e-02$	$7.38e-02$	$7.08e-04$	$1.02e-04$	$3.96e-03$
2	$2.06e-02$	$2.88e-02$	$3.15e-02$	$2.12e-04$	$2.20e-05$	$4.59e-04$
3	$2.11e-04$	$2.76e-03$	$3.12e-03$	$2.42e-04$	$4.68e-06$	$7.90e-05$
4	$3.86e-03$	$1.84e-03$	$2.22e-03$	$6.49e-05$	$8.61e-07$	$1.52e-05$
5	$6.37e-05$	$1.40e-04$	$3.14e-04$	$3.02e-06$	$1.21e-07$	$2.27e-06$
6	$1.34e-04$	$7.94e-05$	$1.32e-04$	$2.75e-06$	$2.83e-08$	$4.37e-07$
7	$1.57e-05$	$2.39e-05$	$3.39e-05$	$-1.76e-08$	$3.97e-09$	$5.49e-08$
8	$5.15e-06$	$1.29e-05$	$1.42e-05$	$1.15e-07$	$1.30e-09$	$1.06e-08$
9	$2.65e-06$	$3.73e-06$	$4.75e-06$	$-3.25e-09$	$1.29e-10$	$1.49e-09$
10	$1.04e-06$	$3.47e-06$	$3.59e-06$	$1.28e-09$	$4.57e-11$	$4.89e-10$
11	$7.24e-07$	$4.15e-07$	$5.59e-07$	$-1.03e-10$	$7.21e-12$	$5.12e-11$
12	$3.47e-07$	$1.24e-07$	$1.46e-07$	$3.32e-11$	$3.19e-12$	$1.26e-11$
13	$1.02e-07$	$4.12e-08$	$6.08e-08$	$-3.92e-12$	$9.62e-13$	$2.30e-12$

Table A.4.7: Example 4. Reliability related consistency error $e^c_{\ell,rel}$ with heuristic estimates $E^{c,1}_{\ell,rel}$, $E^{c,2}_{\ell,rel}$, $E^{c,3}_{\ell,rel}$ and efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ with heuristic estimate $E^c_{\ell,eff}$ for uniform refinement.

ℓ	$e^c_{\ell,rel}$	$E^{c,1}_{\ell,rel}$	$E^{c,2}_{\ell,rel}$	$E^{c,3}_{\ell,rel}$	$e^{c,L^2}_{\ell,eff}$	$E^c_{\ell,eff}$
0	$-2.47e-03$	$1.02e-01$	$1.45e-01$	$-1.21e-03$	$7.82e-04$	$6.35e-04$
1	$-2.53e-03$	$6.42e-02$	$6.42e-02$	$-1.60e-03$	$7.82e-04$	$3.54e-04$
2	$-8.81e-05$	$8.83e-03$	$2.31e-02$	$-1.72e-04$	$3.93e-04$	$8.94e-05$
3	$6.00e-04$	$4.00e-04$	$3.67e-03$	$2.09e-05$	$6.68e-05$	$1.34e-04$
4	$1.05e-04$	$3.17e-04$	$1.66e-03$	$2.49e-06$	$2.00e-05$	$4.27e-05$
5	$2.00e-04$	$2.87e-04$	$3.84e-04$	$9.84e-07$	$1.26e-05$	$1.89e-05$
6	$8.14e-05$	$2.20e-04$	$2.94e-04$	$1.25e-06$	$2.94e-07$	$3.00e-06$
7	$1.12e-05$	$3.51e-05$	$5.53e-05$	$2.13e-07$	$2.49e-07$	$7.45e-07$
8	$1.04e-05$	$5.61e-05$	$7.23e-05$	$-1.24e-07$	$2.49e-07$	$5.36e-07$
9	$7.85e-06$	$1.52e-05$	$1.53e-05$	$-1.05e-08$	$2.02e-07$	$3.85e-07$
10	$2.85e-06$	$1.04e-05$	$1.05e-05$	$-2.85e-08$	$1.91e-07$	$3.72e-07$
11	$2.04e-06$	$4.99e-06$	$5.25e-06$	$-4.57e-09$	$3.99e-08$	$1.34e-07$
12	$1.22e-06$	$2.39e-06$	$2.50e-06$	$-3.19e-09$	$1.89e-08$	$1.08e-07$
13	$6.77e-07$	$1.10e-06$	$1.17e-06$	$-9.54e-10$	$1.94e-09$	$4.35e-09$
14	$4.21e-07$	$9.43e-07$	$9.80e-07$	$-6.73e-10$	$1.75e-09$	$3.69e-09$

Table A.4.8: Example 4. Reliability related consistency error $e^c_{\ell,rel}$ with heuristic estimates $E^{c,1}_{\ell,rel}$, $E^{c,2}_{\ell,rel}$, $E^{c,3}_{\ell,rel}$ and efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ with heuristic estimate $E^c_{\ell,eff}$ for adaptive refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$	$\|\|\|e_\ell\|\|\|$	$\|\|e_{\ell,u}\|\|_{0,\Omega}$	$\|\|e_{\ell,y}\|\|_{1,\Omega}$	$\|\|e_{\ell,p}\|\|_{1,\Omega}$
1	0.40	0.40	0.40	0.24	0.65	0.24	0.24
2	0.29	0.29	0.29	0.34	0.69	0.34	0.34
3	0.44	0.45	0.44	0.61	1.20	0.61	0.61
4	0.39	0.39	0.39	0.39	0.78	0.39	0.39
5	0.52	0.52	0.52	0.57	1.14	0.57	0.57
6	0.42	0.42	0.42	0.41	0.81	0.41	0.41
7	0.54	0.54	0.54	0.57	1.15	0.57	0.57
8	0.44	0.44	0.44	0.42	0.83	0.42	0.42
9	0.55	0.55	0.55	0.57	1.15	0.57	0.57
10	0.44	0.44	0.44	0.42	0.84	0.42	0.42
11	0.55	0.55	0.55	0.57	1.15	0.57	0.57
12	0.45	0.45	0.45	0.42	0.84	0.42	0.42
13	0.55	0.55	0.55	0.57	1.15	0.57	0.57

Table A.4.9: Example 4. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, $\eta_\ell(p)$, $|||e_\ell|||$, $||e_{\ell,u}||_{0,\Omega}$, $||e_{\ell,y}||_{1,\Omega}$, and $||e_{\ell,p}||_{1,\Omega}$ for uniform refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$	$\|\|\|e_\ell\|\|\|$	$\|\|e_{\ell,u}\|\|_{0,\Omega}$	$\|\|e_{\ell,y}\|\|_{1,\Omega}$	$\|\|e_{\ell,p}\|\|_{1,\Omega}$
1	0.69	0.68	0.69	0.54	0.60	0.54	0.54
2	0.49	0.49	0.49	0.61	1.42	0.61	0.61
3	0.51	0.51	0.51	0.63	1.33	0.63	0.63
4	0.53	0.53	0.53	0.47	0.95	0.47	0.47
5	0.49	0.49	0.49	0.58	1.16	0.58	0.58
6	0.53	0.53	0.53	0.47	0.88	0.47	0.47
7	0.48	0.48	0.48	0.53	1.12	0.53	0.53
8	0.54	0.54	0.54	0.49	0.90	0.49	0.49
9	0.47	0.47	0.47	0.52	1.07	0.52	0.52
10	0.52	0.52	0.52	0.47	0.85	0.47	0.47
11	0.49	0.49	0.49	0.53	1.12	0.53	0.53
12	0.51	0.51	0.51	0.47	0.88	0.47	0.47
13	0.49	0.49	0.49	0.52	1.08	0.52	0.52
14	0.50	0.50	0.50	0.48	0.88	0.48	0.48

Table A.4.10: Example 4. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, $\eta_\ell(p)$, $|||e_\ell|||$, $||e_{\ell,u}||_{0,\Omega}$, $||e_{\ell,y}||_{1,\Omega}$, and $||e_{\ell,p}||_{1,\Omega}$ for adaptive refinement.

ℓ	N_ℓ^I	N_ℓ^A	N_ℓ^Z	N_ℓ^C	k'_ℓ	t_ℓ
0	23	2	24	1	2	20”
1	28	21	36	13	6	16”
2	48	65	78	35	6	32”
3	97	128	152	73	7	60”
4	189	292	336	145	8	2’ 02”
5	366	595	656	305	8	4’ 35”
6	741	1244	1371	614	10	7’ 57”
7	1468	2501	2741	1228	9	18’ 60”
8	2984	5081	5613	2452	14	31’ 17”
9	5978	10151	11244	4885	17	75’ 38”
10	12069	20444	22766	9747	20	124’ 55”
11	24256	40769	45587	19438	31	300’ 45”
12	48952	81609	91820	38741	35	498’ 53”
13	98442	162679	183662	77459	60	1203’ 09”

Table A.4.11: Example 4. Number of inactive N_ℓ^I, active N_ℓ^A, zero N_ℓ^Z, and strongly active nodal points N_ℓ^C, number of iterations of PDASS(LSC,ϵ) k'_ℓ and computational time t_ℓ in minutes and seconds for uniform refinement.

ℓ	N_ℓ^I	N_ℓ^A	N_ℓ^Z	N_ℓ^C	k'_ℓ	t_ℓ
0	23	2	24	1	2	20”
1	37	2	38	1	2	37”
2	68	10	71	7	3	1’ 19”
3	137	14	141	10	5	1’ 55”
4	280	29	296	13	6	4’ 34”
5	606	49	638	17	7	9’ 08”
6	1138	65	1179	24	6	17’ 49”
7	2336	103	2414	25	8	31’ 36”
8	4394	116	4483	27	8	67’ 18”
9	8902	227	9088	41	10	120’ 02”
10	16596	255	16812	39	12	262’ 00”
11	33770	389	34110	49	13	452’ 17”
12	63222	520	63684	58	15	976’ 14”
13	127312	754	128005	61	20	1750’ 36”
14	241613	865	242412	66	23	3767’ 39”

Table A.4.12: Example 4. Number of inactive N_ℓ^I, active N_ℓ^A, zero N_ℓ^Z, and strongly active nodal points N_ℓ^C, number of iterations of PDASS(LSC,ϵ) k'_ℓ and computational time t_ℓ in minutes and seconds for adaptive refinement.

A.5 Example 5

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$2.09e+00$	$1.50e+00$	$1.46e+00$
1	$1.88e+00$	$1.34e+00$	$1.31e+00$
2	$1.54e+00$	$1.10e+00$	$1.08e+00$
3	$1.30e+00$	$9.35e-01$	$9.07e-01$
4	$1.05e+00$	$7.48e-01$	$7.40e-01$
5	$8.11e-01$	$5.76e-01$	$5.71e-01$
6	$6.10e-01$	$4.33e-01$	$4.30e-01$
7	$4.52e-01$	$3.21e-01$	$3.19e-01$
8	$3.38e-01$	$2.40e-01$	$2.38e-01$
9	$2.51e-01$	$1.78e-01$	$1.77e-01$
10	$1.88e-01$	$1.33e-01$	$1.33e-01$
11	$1.41e-01$	$9.99e-02$	$9.95e-02$
12	$1.06e-01$	$7.54e-02$	$7.51e-02$

Table A.5.1: Example 5. Error estimator η_ℓ and its components $\eta_\ell(y)$ and $\eta_\ell(p)$ for uniform refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$2.09e+00$	$1.50e+00$	$1.46e+00$
1	$1.67e+00$	$1.19e+00$	$1.17e+00$
2	$1.29e+00$	$9.18e-01$	$9.00e-01$
3	$8.69e-01$	$6.17e-01$	$6.12e-01$
4	$5.81e-01$	$4.12e-01$	$4.10e-01$
5	$4.15e-01$	$2.94e-01$	$2.92e-01$
6	$2.82e-01$	$2.01e-01$	$1.98e-01$
7	$1.98e-01$	$1.41e-01$	$1.39e-01$
8	$1.39e-01$	$9.92e-02$	$9.79e-02$
9	$9.90e-02$	$7.05e-02$	$6.96e-02$
10	$7.07e-02$	$5.04e-02$	$4.97e-02$
11	$5.14e-02$	$3.66e-02$	$3.62e-02$
12	$3.68e-02$	$2.62e-02$	$2.58e-02$
13	$2.66e-02$	$1.89e-02$	$1.87e-02$
14	$1.91e-02$	$1.36e-02$	$1.34e-02$

Table A.5.2: Example 5. Error estimator η_ℓ and its components $\eta_\ell(y)$ and $\eta_\ell(p)$ for adaptive refinement.

ℓ	DOF	$\|\|\|e_\ell\|\|\|$	$\|e_{\ell,u}\|_{0,\Omega}$	$\|e_{\ell,y}\|_{1,\Omega}$	$\|e_{\ell,p}\|_{1,\Omega}$
0	33	$8.09e-01$	$7.71e-02$	$5.65e-01$	$5.74e-01$
1	80	$7.25e-01$	$6.76e-02$	$5.11e-01$	$5.11e-01$
2	161	$6.17e-01$	$4.96e-02$	$4.33e-01$	$4.37e-01$
3	352	$4.38e-01$	$2.34e-02$	$3.09e-01$	$3.09e-01$
4	705	$3.25e-01$	$1.28e-02$	$2.30e-01$	$2.30e-01$
5	1472	$2.34e-01$	$6.86e-03$	$1.65e-01$	$1.65e-01$
6	2945	$1.72e-01$	$3.87e-03$	$1.21e-01$	$1.21e-01$
7	6016	$1.28e-01$	$2.17e-03$	$9.07e-02$	$9.07e-02$
8	12033	$9.58e-02$	$1.28e-03$	$6.77e-02$	$6.77e-02$
9	24320	$7.24e-02$	$7.40e-04$	$5.12e-02$	$5.12e-02$
10	48641	$5.49e-02$	$4.47e-04$	$3.88e-02$	$3.88e-02$
11	97792	$4.19e-02$	$2.66e-04$	$2.97e-02$	$2.97e-02$
12	195585	$3.22e-02$	$1.62e-04$	$2.28e-02$	$2.28e-02$

Table A.5.3: Example 5. Number of DOF, total error $|||e_\ell|||$ and its components $\|e_{\ell,u}\|_{0,\Omega}$, $\|e_{\ell,y}\|_{1,\Omega}$, and $\|e_{\ell,p}\|_{1,\Omega}$ for uniform refinement.

ℓ	DOF	$\|\|\|e_\ell\|\|\|$	$\|e_{\ell,u}\|_{0,\Omega}$	$\|e_{\ell,y}\|_{1,\Omega}$	$\|e_{\ell,p}\|_{1,\Omega}$
0	33	$8.09e-01$	$7.71e-02$	$5.65e-01$	$5.74e-01$
1	42	$6.11e-01$	$4.26e-02$	$4.31e-01$	$4.31e-01$
2	65	$4.02e-01$	$2.06e-02$	$2.83e-01$	$2.84e-01$
3	127	$2.46e-01$	$8.91e-03$	$1.74e-01$	$1.74e-01$
4	280	$1.57e-01$	$3.36e-03$	$1.11e-01$	$1.11e-01$
5	487	$1.07e-01$	$1.47e-03$	$7.57e-02$	$7.57e-02$
6	992	$7.19e-02$	$6.59e-04$	$5.08e-02$	$5.08e-02$
7	1913	$5.01e-02$	$3.19e-04$	$3.54e-02$	$3.55e-02$
8	3653	$3.49e-02$	$1.52e-04$	$2.47e-02$	$2.47e-02$
9	7135	$2.47e-02$	$7.43e-05$	$1.75e-02$	$1.75e-02$
10	13723	$1.75e-02$	$3.83e-05$	$1.24e-02$	$1.24e-02$
11	25987	$1.27e-02$	$7.17e-05$	$8.98e-03$	$8.98e-03$
12	49888	$9.00e-03$	$3.43e-05$	$6.37e-03$	$6.37e-03$
13	95299	$6.51e-03$	$6.77e-06$	$4.60e-03$	$4.60e-03$
14	182618	$4.65e-03$	$4.42e-06$	$3.29e-03$	$3.29e-03$

Table A.5.4: Example 5. Number of DOF, total error $|||e_\ell|||$ and its components $\|e_{\ell,u}\|_{0,\Omega}$, $\|e_{\ell,y}\|_{1,\Omega}$, and $\|e_{\ell,p}\|_{1,\Omega}$ for adaptive refinement.

ℓ	$osc_{\ell,eff}$	$osc_{\mathcal{Z}_\ell}(f)$	$osc_{\mathcal{I}_\ell}(y^d)$	$osc_{\ell,rel}$
0	$4.92e-01$	$3.52e-01$	$3.44e-01$	$0.00e+00$
1	$3.28e-01$	$2.38e-01$	$2.25e-01$	$0.00e+00$
2	$1.15e-01$	$8.22e-02$	$8.10e-02$	$0.00e+00$
3	$5.35e-02$	$3.85e-02$	$3.72e-02$	$0.00e+00$
4	$2.28e-02$	$1.62e-02$	$1.61e-02$	$0.00e+00$
5	$8.80e-03$	$6.26e-03$	$6.19e-03$	$0.00e+00$
6	$3.80e-03$	$2.69e-03$	$2.69e-03$	$0.00e+00$
7	$1.51e-03$	$1.10e-03$	$1.03e-03$	$0.00e+00$
8	$6.32e-04$	$4.70e-04$	$4.22e-04$	$0.00e+00$
9	$2.61e-04$	$1.90e-04$	$1.79e-04$	$0.00e+00$
10	$1.13e-04$	$8.08e-05$	$7.98e-05$	$0.00e+00$
11	$4.66e-05$	$3.41e-05$	$3.18e-05$	$0.00e+00$
12	$1.87e-05$	$1.43e-05$	$1.20e-05$	$0.00e+00$

Table A.5.5: Example 5. Efficiency related data oscillation $osc_{\ell,eff}$, its components $osc_{\mathcal{Z}_\ell}(f)$, $osc_{\mathcal{I}_\ell}(y^d)$, and reliability related data oscillation $osc_{\ell,rel}$ for uniform refinement.

ℓ	$osc_{\ell,eff}$	$osc_{\mathcal{Z}_\ell}(f)$	$osc_{\mathcal{I}_\ell}(y^d)$	$osc_{\ell,rel}$
0	$4.92e-01$	$3.52e-01$	$3.44e-01$	$0.00e+00$
1	$2.23e-01$	$1.73e-01$	$1.40e-01$	$0.00e+00$
2	$8.00e-02$	$5.93e-02$	$5.38e-02$	$0.00e+00$
3	$2.42e-02$	$1.86e-02$	$1.55e-02$	$0.00e+00$
4	$1.34e-02$	$1.27e-02$	$4.44e-03$	$0.00e+00$
5	$1.20e-02$	$1.19e-02$	$1.19e-03$	$0.00e+00$
6	$1.18e-02$	$1.18e-02$	$3.68e-04$	$0.00e+00$
7	$1.18e-02$	$1.18e-02$	$1.32e-04$	$0.00e+00$
8	$1.18e-02$	$1.18e-02$	$5.44e-05$	$0.00e+00$
9	$7.14e-03$	$7.14e-03$	$2.26e-05$	$0.00e+00$
10	$9.96e-03$	$9.96e-03$	$6.03e-06$	$0.00e+00$
11	$1.72e-03$	$1.72e-03$	$3.27e-06$	$0.00e+00$
12	$2.12e-03$	$2.12e-03$	$1.22e-06$	$0.00e+00$
13	$7.75e-04$	$7.75e-04$	$3.47e-07$	$0.00e+00$
14	$3.13e-04$	$3.13e-04$	$9.19e-08$	$0.00e+00$

Table A.5.6: Example 5. Efficiency related data oscillation $osc_{\ell,eff}$, its components $osc_{\mathcal{Z}_\ell}(f)$, $osc_{\mathcal{I}_\ell}(y^d)$, and reliability related data oscillation $osc_{\ell,rel}$ for adaptive refinement.

ℓ	$e^c_{\ell,rel}$	$E^{c,1}_{\ell,rel}$	$E^{c,2}_{\ell,rel}$	$E^{c,3}_{\ell,rel}$	$e^{c,L^2}_{\ell,eff}$	$E^c_{\ell,eff}$
0	$-8.02e-05$	$2.21e-03$	$-1.48e-02$	$-3.61e-02$	$3.00e-02$	$1.47e-02$
1	$4.87e-02$	$2.32e-03$	$-8.05e-03$	$-1.74e-02$	$1.63e-03$	$2.16e-02$
2	$-4.16e-06$	$1.85e-02$	$-5.29e-02$	$-1.00e-01$	$3.74e-03$	$3.18e-03$
3	$4.57e-04$	$3.68e-02$	$-2.63e-02$	$-1.11e-01$	$1.90e-04$	$3.77e-04$
4	$2.29e-04$	$4.25e-02$	$-3.90e-02$	$-9.79e-02$	$9.23e-05$	$2.72e-04$
5	$2.22e-04$	$3.41e-02$	$-2.73e-02$	$-8.69e-02$	$1.06e-05$	$2.24e-04$
6	$2.04e-04$	$2.58e-02$	$-2.44e-02$	$-5.83e-02$	$1.47e-06$	$1.07e-04$
7	$1.61e-04$	$2.05e-02$	$-1.95e-02$	$-4.95e-02$	$7.77e-07$	$3.89e-05$
8	$3.96e-05$	$1.51e-02$	$-1.46e-02$	$-3.08e-02$	$3.36e-07$	$2.47e-05$
9	$1.91e-05$	$1.16e-02$	$-1.14e-02$	$-2.62e-02$	$1.40e-07$	$2.22e-06$
10	$4.92e-07$	$8.30e-03$	$-8.25e-03$	$-1.58e-02$	$3.31e-08$	$5.24e-07$
11	$6.43e-07$	$6.07e-03$	$-6.05e-03$	$-1.36e-02$	$1.82e-08$	$2.08e-07$
12	$1.11e-07$	$4.30e-03$	$-4.29e-03$	$-8.06e-03$	$5.83e-09$	$1.10e-07$

Table A.5.7: Example 5. Reliability related consistency error $e^c_{\ell,rel}$ with heuristic estimates $E^{c,1}_{\ell,rel}$, $E^{c,2}_{\ell,rel}$, $E^{c,3}_{\ell,rel}$ and efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ with heuristic estimate $E^c_{\ell,eff}$ for uniform refinement.

ℓ	$e^c_{\ell,rel}$	$E^{c,1}_{\ell,rel}$	$E^{c,2}_{\ell,rel}$	$E^{c,3}_{\ell,rel}$	$e^{c,L^2}_{\ell,eff}$	$E^c_{\ell,eff}$
0	$-8.02e-05$	$2.21e-03$	$-1.48e-02$	$-3.61e-02$	$3.00e-02$	$1.47e-02$
1	$-2.41e-06$	$2.03e-02$	$-4.94e-02$	$-9.19e-02$	$1.79e-03$	$5.55e-04$
2	$-1.30e-06$	$6.51e-02$	$-5.90e-02$	$-1.13e-01$	$1.79e-03$	$5.01e-04$
3	$1.02e-20$	$5.66e-02$	$-5.37e-02$	$-6.99e-02$	$0.00e+00$	$4.20e-04$
4	$3.50e-04$	$2.78e-02$	$-2.24e-02$	$-4.19e-02$	$0.00e+00$	$2.45e-04$
5	$2.72e-05$	$1.71e-02$	$-1.61e-02$	$-2.70e-02$	$0.00e+00$	$3.76e-05$
6	$7.29e-07$	$1.10e-02$	$-1.08e-02$	$-1.76e-02$	$0.00e+00$	$4.29e-07$
7	$3.59e-06$	$6.89e-03$	$-6.66e-03$	$-1.06e-02$	$0.00e+00$	$1.32e-06$
8	$4.26e-07$	$4.63e-03$	$-4.55e-03$	$-7.05e-03$	$0.00e+00$	$4.72e-07$
9	$3.22e-07$	$3.16e-03$	$-3.12e-03$	$-4.89e-03$	$0.00e+00$	$3.50e-07$
10	$8.09e-08$	$2.24e-03$	$-2.22e-03$	$-3.50e-03$	$1.63e-03$	$1.32e-03$
11	$9.29e-10$	$1.54e-03$	$-1.54e-03$	$-2.44e-03$	$1.99e-04$	$6.39e-04$
12	$2.50e-09$	$1.12e-03$	$-1.12e-03$	$-1.75e-03$	$3.18e-05$	$8.75e-04$
13	$5.33e-09$	$7.89e-04$	$-7.88e-04$	$-1.23e-03$	$3.65e-05$	$1.19e-04$
14	$1.16e-09$	$5.71e-04$	$-5.70e-04$	$-8.75e-04$	$1.27e-05$	$5.31e-05$

Table A.5.8: Example 5. Reliability related consistency error $e^c_{\ell,rel}$ with heuristic estimates $E^{c,1}_{\ell,rel}$, $E^{c,2}_{\ell,rel}$, $E^{c,3}_{\ell,rel}$ and efficiency related consistency error $e^{c,L^2}_{\ell,eff}$ with heuristic estimate $E^c_{\ell,eff}$ for adaptive refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$	$\|\|\|e_\ell\|\|\|$	$\|\|e_{\ell,u}\|\|_{0,\Omega}$	$\|\|e_{\ell,y}\|\|_{1,\Omega}$	$\|\|e_{\ell,p}\|\|_{1,\Omega}$
1	0.12	0.13	0.12	0.12	0.15	0.11	0.13
2	0.28	0.28	0.28	0.23	0.44	0.24	0.22
3	0.21	0.21	0.22	0.44	0.96	0.43	0.44
4	0.31	0.32	0.29	0.43	0.87	0.43	0.43
5	0.35	0.35	0.35	0.45	0.84	0.45	0.45
6	0.41	0.41	0.41	0.45	0.83	0.45	0.45
7	0.42	0.42	0.42	0.41	0.81	0.41	0.41
8	0.42	0.42	0.42	0.42	0.77	0.42	0.42
9	0.42	0.42	0.42	0.40	0.77	0.40	0.40
10	0.42	0.42	0.42	0.40	0.73	0.40	0.40
11	0.41	0.41	0.41	0.39	0.74	0.39	0.39
12	0.41	0.41	0.41	0.38	0.72	0.38	0.38

Table A.5.9: Example 5. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, $\eta_\ell(p)$, $|||e_\ell|||$, $||e_{\ell,u}||_{0,\Omega}$, $||e_{\ell,y}||_{1,\Omega}$, and $||e_{\ell,p}||_{1,\Omega}$ for uniform refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$	$\|\|\|e_\ell\|\|\|$	$\|\|e_{\ell,u}\|\|_{0,\Omega}$	$\|\|e_{\ell,y}\|\|_{1,\Omega}$	$\|\|e_{\ell,p}\|\|_{1,\Omega}$
1	0.93	0.96	0.91	1.16	2.46	1.12	1.18
2	0.60	0.60	0.60	0.96	1.66	0.96	0.95
3	0.58	0.59	0.58	0.73	1.25	0.73	0.73
4	0.51	0.51	0.51	0.57	1.23	0.57	0.57
5	0.61	0.61	0.61	0.69	1.50	0.69	0.69
6	0.54	0.54	0.54	0.56	1.13	0.56	0.56
7	0.54	0.54	0.54	0.55	1.10	0.55	0.55
8	0.54	0.54	0.54	0.56	1.14	0.56	0.56
9	0.51	0.51	0.51	0.52	1.07	0.51	0.52
10	0.51	0.51	0.52	0.53	1.01	0.53	0.53
11	0.50	0.50	0.50	0.50	−0.98	0.50	0.50
12	0.51	0.51	0.52	0.53	1.13	0.53	0.53
13	0.50	0.50	0.50	0.50	2.51	0.50	0.50
14	0.51	0.51	0.51	0.52	0.65	0.52	0.52

Table A.5.10: Example 5. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, $\eta_\ell(p)$, $|||e_\ell|||$, $||e_{\ell,u}||_{0,\Omega}$, $||e_{\ell,y}||_{1,\Omega}$, and $||e_{\ell,p}||_{1,\Omega}$ for adaptive refinement.

ℓ	N_ℓ^I	N_ℓ^A	N_ℓ^Z	N_ℓ^C	k'_ℓ	t_ℓ
0	14	19	20	13	152	17”
1	3	77	43	37	3	13”
2	70	91	115	46	5	1’ 12”
3	42	310	250	102	4	1’ 27”
4	92	613	528	177	4	2’ 50”
5	95	1377	1072	400	5	5’ 32”
6	223	2722	2220	725	6	11’ 28”
7	325	5691	4580	1436	7	23’ 34”
8	581	11452	9279	2754	7	45’ 03”
9	1223	23097	18819	5501	8	96’ 10”
10	3318	45323	37904	10737	10	228’ 49”
11	5266	92526	76254	21538	11	407’ 20”
12	9472	186113	152921	42664	15	763’ 36”

Table A.5.11: Example 5. Number of inactive N_ℓ^I, active N_ℓ^A, zero N_ℓ^Z, and strongly active nodal points N_ℓ^C, number of iterations of PDASS(LSC,ϵ) k'_ℓ and computational time t_ℓ in minutes and seconds for uniform refinement.

ℓ	N_ℓ^I	N_ℓ^A	N_ℓ^Z	N_ℓ^C	k'_ℓ	t_ℓ
0	14	19	20	13	152	17”
1	10	32	21	21	5	13”
2	27	38	48	17	5	46”
3	90	37	108	19	5	1’ 22”
4	204	76	255	25	6	2’ 17”
5	353	134	463	24	7	6’ 44”
6	749	243	969	23	10	10’ 08”
7	1576	337	1875	38	12	20’ 43”
8	3198	455	3605	48	15	37’ 15”
9	6446	689	7051	84	15	65’ 10”
10	12682	1041	13628	95	21	126’ 06”
11	24537	1450	25924	63	152	512’ 14”
12	47790	2098	49794	94	36	550’ 47”
13	91943	3356	95122	177	37	885’ 17”
14	177918	4700	182373	245	68	1840’ 14”

Table A.5.12: Example 5. Number of inactive N_ℓ^I, active N_ℓ^A, zero N_ℓ^Z, and strongly active nodal points N_ℓ^C, number of iterations of PDASS(LSC,ϵ) k'_ℓ and computational time t_ℓ in minutes and seconds for adaptive refinement.

A.6 Example 6

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$5.56e-02$	$1.18e-02$	$5.44e-02$
1	$4.90e-02$	$6.93e-03$	$4.85e-02$
2	$4.08e-02$	$5.10e-03$	$4.04e-02$
3	$3.03e-02$	$3.58e-03$	$3.01e-02$
4	$2.29e-02$	$2.58e-03$	$2.27e-02$
5	$1.77e-02$	$1.86e-03$	$1.76e-02$
6	$1.29e-02$	$1.34e-03$	$1.29e-02$
7	$1.02e-02$	$9.81e-04$	$1.02e-02$
8	$7.42e-03$	$7.05e-04$	$7.39e-03$
9	$5.94e-03$	$5.30e-04$	$5.92e-03$

Table A.6.1: Example 6. Error estimator η_ℓ and its components $\eta_\ell(y)$ and $\eta_\ell(p)$ for uniform refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
0	$5.56e-02$	$1.18e-02$	$5.44e-02$
1	$4.22e-02$	$1.06e-02$	$4.09e-02$
2	$4.92e-02$	$8.24e-03$	$4.85e-02$
3	$4.19e-02$	$7.20e-03$	$4.13e-02$
4	$3.03e-02$	$6.21e-03$	$2.97e-02$
5	$2.27e-02$	$4.37e-03$	$2.23e-02$
6	$1.63e-02$	$3.67e-03$	$1.59e-02$
7	$1.31e-02$	$2.55e-03$	$1.29e-02$
8	$9.98e-03$	$2.10e-03$	$9.75e-03$
9	$7.73e-03$	$1.48e-03$	$7.59e-03$
10	$5.59e-03$	$1.15e-03$	$5.47e-03$
11	$4.22e-03$	$8.31e-04$	$4.14e-03$
12	$3.16e-03$	$6.21e-04$	$3.10e-03$
13	$2.40e-03$	$4.66e-04$	$2.35e-03$

Table A.6.2: Example 6. Error estimator η_ℓ and its components $\eta_\ell(y)$ and $\eta_\ell(p)$ for adaptive refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
1	0.17	0.70	0.15
2	0.23	0.39	0.23
3	0.40	0.47	0.39
4	0.37	0.43	0.37
5	0.35	0.45	0.34
6	0.43	0.45	0.43
7	0.33	0.43	0.32
8	0.44	0.45	0.44
9	0.31	0.40	0.31

Table A.6.3: Example 6. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, and $\eta_\ell(p)$ for uniform refinement.

ℓ	η_ℓ	$\eta_\ell(y)$	$\eta_\ell(p)$
1	0.78	0.28	0.81
2	−0.29	0.49	−0.33
3	0.41	0.34	0.41
4	0.92	0.42	0.93
5	0.45	0.55	0.44
6	0.55	0.29	0.56
7	0.32	0.54	0.31
8	0.50	0.36	0.51
9	0.43	0.59	0.43
10	0.50	0.39	0.51
11	0.49	0.57	0.48
12	0.44	0.44	0.44
13	0.49	0.50	0.49

Table A.6.4: Example 6. Experimental convergence rates associated with η_ℓ, $\eta_\ell(y)$, and $\eta_\ell(p)$ for adaptive refinement.

ℓ	N_ℓ^I	N_ℓ^A	N_ℓ^Z	N_ℓ^C	k'_ℓ	t_ℓ
0	79	118	161	36	5	2' 03"
1	228	189	343	74	153	3' 48"
2	574	344	751	167	152	7' 15"
3	1328	618	1615	331	153	10' 19"
4	3012	1148	3460	700	153	25' 49"
5	6527	2149	7261	1415	153	46' 31"
6	14092	4025	15219	2898	153	111' 45"
7	29707	7700	31569	5838	202	227' 01"
8	62566	14684	65439	11811	152	472' 04"
9	129841	28409	134538	23712	152	954' 23"

Table A.6.5: Example 6. Number of inactive N_ℓ^I, active N_ℓ^A, zero N_ℓ^Z, and strongly active nodal points N_ℓ^C, number of iterations of PDASS(LSC,ϵ) k'_ℓ and computational time t_ℓ in minutes and seconds for uniform refinement.

ℓ	N_ℓ^I	N_ℓ^A	N_ℓ^Z	N_ℓ^C	k'_ℓ	t_ℓ
0	79	118	161	36	5	2' 03"
1	154	126	244	36	5	5' 35"
2	344	130	437	37	152	9' 20"
3	535	167	662	40	202	16' 04"
4	784	215	952	47	152	20' 49"
5	1660	242	1854	48	152	41' 34"
6	3153	328	3426	55	153	77' 36"
7	6448	398	6755	91	153	138' 38"
8	11239	534	11668	105	19	280' 31"
9	20427	751	21021	157	202	408' 34"
10	39456	976	40251	181	152	890' 35"
11	70277	1486	71435	328	43	1382' 04"
12	137461	1886	138953	394	61	2777' 23"
13	243504	3099	246027	576	153	4976' 22"

Table A.6.6: Example 6. Number of inactive N_ℓ^I, active N_ℓ^A, zero N_ℓ^Z, and strongly active nodal points N_ℓ^C, number of iterations of PDASS(LSC,ϵ) k'_ℓ and computational time t_ℓ in minutes and seconds for adaptive refinement.

B List of abbreviations

cf.	confer (compare)
i.e.	id est (that is)
e.g.	exempli gratia (for example)
w.r.t.	with respect to
a.e.	almost everywhere

Bibliography

[1] Robert A. Adams and John J. F. Fournier, *Sobolev spaces. 2nd ed.*, Pure and Applied Mathematics 140. New York, NY: Academic Press. xiii, 305 p. , 2003 (English).

[2] Mark Ainsworth, J.Tinsley Oden, and C.Y. Lee, *Local a posteriori error estimators for variational inequalities.*, Numer. Methods Partial Differ. Equations **9** (1993), no. 1, 23–33 (English).

[3] Hans Wilhelm Alt, *Linear functional analysis. An application oriented introduction. (Lineare Funktionalanalysis. Eine anwendungsorientierte Einführung.) 5th revised ed.*, Berlin: Springer. xiv, 431 p. , 2006 (German).

[4] Ivo Babuška and Theofanis Strouboulis, *The finite element method and its reliability.*, Oxford: Clarendon Press. xii , 2001 (English).

[5] V. Barbu, *Optimal control of variational inequalities.*, Research Notes in Mathematics, 100. Boston - London - Melbourne: Pitman Advanced Publishing Program. 298 p., 1984 (English).

[6] M. Bergounioux, M. Haddou, M. Hintermüller, and K. Kunisch, *A comparison of a Moreau-Yosida-based active set strategy and interior point methods for constrained optimal control problems.*, SIAM J. Optim. **11** (2000), no. 2, 495–521 (English).

[7] Maïtine Bergounioux, *Optimal control of an obstacle problem.*, Appl. Math. Optim. **36** (1997), 147–172 (English).

[8] ———, *Optimal control of problems governed by abstract elliptic variational inequalities with state constraints.*, SIAM J. Control Optimization **36** (1998), no. 1, 273–289 (English).

[9] ———, *Optimal control of semilinear elliptic obstacle problems.*, J. Nonlinear Convex Anal. **3** (2002), no. 1, 25–39 (English).

[10] Maïtine Bergounioux and Fulbert Mignot, *Optimal control of obstacle problems: Existence of Lagrange multipliers.*, ESAIM, Control Optim. Calc. Var. **5** (2000), 45–70 (English).

[11] Dietrich Braess, *A posteriori error estimators for obstacle problems – another look.*, Numer. Math. **101** (2005), no. 3, 415–421 (English).

[12] Dietrich Braess, Carsten Carstensen, and Ronald H.W. Hoppe, *Convergence analysis of a conforming adaptive finite element method for an obstacle problem.*, Numer. Math. **107** (2007), no. 3, 455–471 (English).

[13] ———, *Error reduction in adaptive finite element approximations of elliptic obstacle problems.*, J. Computational Math. **27** (2009), no. 2-3, 148–169 (English).

[14] Dietrich Braess, Ronald H.W. Hoppe, and Joachim Schöberl, *A posteriori estimators for obstacle problems by the hypercircle method.*, Comp. Visual. Sci. **11** (2008), no. 4-6, 351–362 (English).

[15] James H. Bramble and Jinchao Xu, *Some estimates for a weighted L^2 projection.*, Math. Comput. **56** (1991), no. 194, 463–476 (English).

[16] J.Manuel Cascon, Christian Kreuzer, Ricardo H. Nochetto, and Kunibert G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method.*, SIAM J. Numer. Anal. **46** (2008), no. 5, 2524–2550 (English).

[17] Zhiming Chen and Ricardo H. Nochetto, *Residual type a posteriori error estimates for elliptic obstacle problems.*, Numer. Math. **84** (2000), no. 4, 527–548 (English).

[18] Philippe G. Ciarlet, *The finite element method for elliptic problems. 1st repr. (paperback).*, Studies in Mathematics and its Applications, Vol. 4. Amsterdam -New York - Oxford: North-Holland Publishing Company. XX, 530 p., 1980 (English).

[19] Ph. Clement, *Approximation by finite element functions using local regularization.*, (1975) (English).

[20] Willy Dörfler, *A convergent adaptive algorithm for Poisson's equation.*, SIAM J. Numer. Anal. **33** (1996), no. 3, 1106–1124 (English).

[21] Duvaut, G. and Lions, J.L., *Inequalities in mechanics and physics*, Grundlehren der mathematischen Wissenschaften, Springer-Verlag, 1976 (English).

[22] Lawrence C. Evans, *Partial differential equations.*, Graduate Studies in Mathematics. 19. Providence, RI: American Mathematical Society (AMS). xvii, 662 p. , 1998 (English).

[23] Lawrence C. Evans and Ronald F. Gariepy, *Measure theory and fine properties of functions.*, Studies in Advanced Mathematics. Boca Raton: CRC Press. viii, 268 p. , 1992 (English).

[24] Francisco Facchinei and Jong-Shi Pang, *Finite-dimensional variational inequalities and complementarity problems.*, Springer Series in Operations Research. New York, NY: Springer. xxxiii, 624 p., 2003 (English).

[25] A. Gaevskaya, R.H.W. Hoppe, and S. Repin, *Functional approach to a posteriori error estimation for elliptic optimal control problems with distributed control.*, J. Math. Sci., New York **144** (2007), no. 6, 4535–4547 (English. Russian original).

[26] Alexandra Gaevskaya, Ronald H.W. Hoppe, and Sergey Repin, *A posteriori estimates for cost functionals of optimal control problems.*, Bermúdez de Castro, Alfredo (ed.) et al., Numerical mathematics and advanced applications. Proceedings of ENUMATH 2005, the 6th European conference on numerical mathematics and advanced applications, Santiago de Compostela, Spain, July 18–22, 2005. Berlin: Springer. 308-316 (2006)., 2006.

[27] Roland Glowinski, *Numerical methods for nonlinear variational problems.*, Springer Series in Computational Physics. New York etc.: Springer-Verlag. XV, 493 p. , 1984 (English).

[28] Roland Glowinski, Jacques-Louis Lions, and Raymond Tremolieres, *Numerical analysis of variational inequalities. Transl. and rev. ed.*, Studies in Mathematics and its Applications, Vol. 8. Amsterdam, New York, Oxford: North-Holland Publishing Company. XXIX, 776 p. $ 109.75; Dfl. 225.00 , 1981 (English).

[29] P. Grisvard, *Elliptic problems in nonsmooth domains.*, Monographs and Studies in Mathematics, 24. Pitman Advanced Publishing Program. Boston-London-Melbourne: Pitman Publishing Inc. XIV, 410 p. , 1985 (English).

[30] A. Günther and M. Hinze, *A posteriori error control of a state constrained elliptic control problem.*, J. Numer. Math. **16** (2008), no. 4, 307–322 (English).

[31] M. Hintermüller, *An active-set equality constrained Newton solver with feasibility restoration for inverse coefficient problems in elliptic variational inequalities.*, (2008) (English).

[32] M. Hintermüller and R.H.W. Hoppe, *Goal-oriented adaptivity in control constrained optimal control of partial differential equations.*, SIAM J. Control Optim. **47** (2008), no. 4, 1721–1743 (English).

[33] M. Hintermüller and I. Kopacka, *Mathematical programs with complementarity constraints in function space: C- and strong stationarity and a path-following algorithm.*, SIAM J. Optim. **20** (2009), no. 2, 868–902 (English).

[34] Michael Hintermüller and Ronald H.W. Hoppe, *Goal-oriented adaptivity in pointwise state constrained optimal control of partial differential equations.*, SIAM J. Control Optim. **48** (2010), no. 8, 5468–5487 (English).

[35] Michael Hintermüller, Ronald H.W. Hoppe, Yuri Iliash, and Michael Kieweg, *An a posteriori error analysis of adaptive finite element methods for distributed elliptic control problems with control constraints.*, ESAIM, Control Optim. Calc. Var. **14** (2008), no. 3, 540–560 (English).

[36] Michael Hintermüller, Ronald H.W Hoppe, and Caroline. Löbhard, *Dual-weighted goal-oriented adaptive finite elements for optimal control of elliptic variational inequalities.*, submitted to ESAIM: Optimal Control and Calculus of Variations (2012) (English).

[37] R.H.W. Hoppe, Y. Iliash, C. Iyyunni, and N.H. Sweilam, *A posteriori error estimates for adaptive finite element discretizations of boundary control problems.*, J. Numer. Math. **14** (2006), no. 1, 57–82 (English).

[38] R.H.W. Hoppe and M. Kieweg, *A posteriori error estimation of finite element approximations of pointwise state constrained distributed control problems.*, J. Numer. Math. **17** (2009), no. 3, 219–244 (English).

[39] ———, *Adaptive finite element methods for mixed control-state constrained optimal control problems for elliptic boundary value problems* ., Comput. Optim. Appl. **46** (2010), no. 3, 511–533 (English).

[40] R.H.W. Hoppe and R. Kornhuber, *Adaptive multilevel methods for obstacle problems.*, SIAM J. Numer. Anal. **31** (1994), no. 2, 301–323 (English).

[41] K. Ito and K. Kunisch, *Optimal control of elliptic variational inequalities.*, Appl. Math. Optimization **41** (2000), no. 3, 343–364 (English).

[42] Claes Johnson, *Adaptive finite element methods for the obstacle problem.*, Math. Models Methods Appl. Sci. **2** (1992), no. 4, 483–487 (English).

[43] Michael Kieweg, *A posteriori error analysis for distributed elliptic optimal control problems with pointewise state constraints*, Ph.D. thesis, Universität Graz, Augsburg, 2008.

[44] David Kinderlehrer and Guido Stampacchia, *An introduction to variational inequalities and their applications. Reprint of the 1980 original.*, Classics in Applied Mathematics. 31. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM). xx, 313 p. , 2000 (English).

[45] Ian Kopacka, *Mpecs/mpccs in functional space: first order optimality concepts, path-following, and multilevel algorithms*, Ph.D. thesis, Karl-Franzens-Universität Graz, Graz, 2009.

[46] Karl Kunisch and Daniel Wachsmuth, *Sufficient optimality conditions and semismooth Newton methods for optimal control of stationary variational inequalities.*, RICAM Report 2009-04 (2010), 1–41 (English).

[47] Ruo Li, Wenbin Liu, Heping Ma, and Tao Tang, *Adaptive finite element approximation for distributed elliptic optimal control problems.*, SIAM J. Control Optimization **41** (2002), no. 5, 1321–1349 (English).

[48] F. Mignot, *Contrôle dans les inéquations variationelles elliptiques.*, J. Funct. Anal. **22** (1976), 130–185 (French).

[49] F. Mignot and J.P. Puel, *Optimal control in some variational inequalities.*, SIAM J. Control Optimization **22** (1984), 466–476 (English).

[50] Pedro Morin, Ricardo H. Nochetto, and Kunibert G. Siebert, *Data oscillation and convergence of adaptive FEM.*, SIAM J. Numer. Anal. **38** (2000), no. 2, 466–488 (English).

[51] Pekka Neittaanmäki and Sergey Repin, *Reliable methods for computer simulation. Error control and a posteriori estimates.*, Studies in Mathematics and its Applications 33. Amsterdam: Elsevier. x, 305 p., 2004 (English).

[52] Ricardo H. Nochetto, Kunibert G. Siebert, and Andreas Veeser, *Pointwise a posteriori error control for elliptic obstacle problems.*, Numer. Math. **95** (2003), no. 1, 163–195 (English).

[53] ———, *Fully localized a posteriori error estimators and barrier sets for contact problems.*, SIAM J. Numer. Anal. **42** (2005), no. 5, 2118–2135 (English).

[54] Jiří Outrata and Jochem Zowe, *A numerical approach to optimization problems with variational inequality constraints.*, Math. Program. **68** (1995), no. 1(A), 105–130 (English).

[55] J.F. Rodrigues, *Obstacle problems in mathematical physics*, Notas de matemática, North-Holland, 1987.

[56] Holger Scheel and Stefan Scholtes, *Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity.*, Math. Oper. Res. **25** (2000), no. 1, 1–22 (English).

[57] L.Ridgway Scott and Shangyou Zhang, *Finite element interpolation of nonsmooth functions satisfying boundary conditions.*, Math. Comput. **54** (1990), no. 190, 483–493 (English).

[58] Robert S. Strichartz, *A guide to distribution theory and Fourier transforms.*, River Edge, NJ: World Scientific. x, 226 p., 2003 (English).

[59] Fredi Tröltzsch, *Optimal control of partial differential equations. Theory, procedures, and applications. (Optimale Steuerung partieller Differentialgleichungen. Theorie, Verfahren und Anwendungen.)*, Wiesbaden: Vieweg. x, 297 p. , 2005 (German).

[60] Andreas Veeser, *Efficient and reliable a posteriori error estimators for elliptic obstacle problems.*, SIAM J. Numer. Anal. **39** (2001), no. 1, 146–167 (English).

[61] Rüdiger Verfürth, *A review of a posteriori error estimation and adaptive mesh-refinement techniques.*, Wiley-Teubner Series Advances in Numerical Mathematics. Chichester: John Wiley & Sons. Stuttgart: B. G. Teubner. vi, 127 p., 1996 (English).

[62] Boris Vexler and Winnifried Wollner, *Adaptive finite elements for elliptic optimization problems with control constraints.*, SIAM J. Control Optim. **47** (2008), no. 1, 509–534 (English).

[63] Joseph Wloka, *Partielle Differentialgleichungen. Sobolevräume und Randwertaufgaben.*, Mathematische Leitfäden. Stuttgart: B. G. Teubner. 500 S., 1982 (German).

Lebenslauf Alexandra Gaevskaya

Persönliche Daten

Geburtsdatum:	18. August 1981
Geburtsort:	St. Petersburg, Russland
Staatsangehörigkeit:	russisch

Beruflicher Werdegang

Seit 03.2013	**Anwendungsentwicklerin**, Allgemeines Rechenzentrum GmbH (ARZ), Innsbruck, Österreich
08.2005 – 09.2010	**Wissenschaftliche Mitarbeiterin**, Lehrstuhl für angewandte Analysis mit Schwerpunkt Numerik, Universität Augsburg, Deutschland
	Lehre: Grundlagen der numerischen Mathematik, Numerische Methoden für gewöhnliche und partielle Differentialgleichungen, Fehlerschätzverfahren. Übungs-, Vorlesungs- und Seminarleitung, Studierendenbetreuung.
11.2004 – 07.2005	**Wissenschaftliche Hilfskraft**, Lehrstuhl für angewandte Analysis mit Schwerpunkt Numerik, Universität Augsburg, Deutschland
09.2004 – 10.2004	**Wissenschaftliche Hilfskraft**, Lehrstuhl für angewandte Mathematik, Staatliche Polytechnische Universität, Sankt Petersburg, Russland
	Lehre: Grundlagen der numerischen Mathematik. Übungsleitung, Studierendenbetreuung.
Seit 09.2004	**Forschung:** Fehlerschätzverfahren für die Lösungen der mittels Finite-Elemente-Methode diskretisierten mathematischen Modelle (insg. 8 Publikationen)
	Projekte: functional type error estimates for parabolic PDEs; functional type and residual based error estimates for optimally controlled elliptic PDEs with and without control constraints; equilibrated residual error estimates for elliptic PDEs; residual based error estimates for optimally controlled elliptic variational inequalities.

Ausbildung und Abschlüsse

2009 – 2013	Arbeit an Promotion in Mathematik, Universität Augsburg, Deutschland

	Thema: Adaptive finite elements for optimally controlled elliptic variational inequalities of obstacle type
06.2004	**M. Sc. Mathematik**, Staatliche Polytechnische Universität, Sankt Petersburg, Russland
	Fachgebiet: Angewandte Mathematik und Informatik
07.2002	**B. Sc. Mathematik**, Staatliche Polytechnische Universität, Sankt Petersburg, Russland
	Fachgebiet: Angewandte Mathematik und Informatik
06.1998	**Abiturabschluss**, Schule 570, Sankt Petersburg, Russland

Publikationen

[1] A. Gaevskaya, and S. Repin; *A posteriori error estimates for initial-boundary value problem for the heat equation.* In proceeding of 5th All-Russian Seminar on Grid Methods in Boundary Value Problems and Applications, Kazan, 2004, pp. 40–43 (in Russian)

[2] A. Gaevskaya, and S. Repin; *A posteriori error estimates for approximate solutions of linear parabolic problems.* Differential Equations, Vol. 41, No. 7, 2005, pp. 970–983

[3] A. Gaevskaya, R.H.W.Hoppe, Y. Iliash, and M. Kieweg. *Convergence analysis of an adaptive finite element method for distributed control problems with control constraints.* Kunisch, Karl (ed.) et al., Control of coupled partial differential equations. Selected papers originating from the conference on optimal control of coupled systems of partial differential equations, Oberwolfach, Germany, April 17–23, 2005. Basel: Birkhäuser. ISNM. International Series of Numerical Mathe- matics 155, 2007, pp. 47–68

[4] A. Gaevskaya, R.H.W. Hoppe, and S. Repin. *A Posteriori Estimates for Cost Functionals of Optimal Control Problems.* Bermúdez de Castro, Alfredo (ed.) et al., Numerical mathematics and advanced applications. Proceedings of ENUMATH 2005, the 6th European conference on numerical mathematics and advanced applications, Santiago de Compostela, Spain, July 18–22, 2005. Berlin: Springer, 2006, pp. 308–316.

[5] A. Gaevskaya, R.H.W. Hoppe, and S. Repin. *A Posteriori Estimates for Optimal Control Problems with Control Constraints.* Mathematisches Forschungsinstitut Oberwolfach, *Numer. Tech. for Optimization Problems with PDE Constraints*, Report 11, 2006, pp. 606–608

[6] A. Gaevskaya. *A Posteriori Estimates for Cost Functionals of the Optimal Control Problems Governed by Elliptic PDEs with Boundary Control.* NTV SPbGTU, 3, 2006, pp. 176–179. (in Russian)

[7] A. Gaevskaya, R.H.W. Hoppe, and S. Repin. *Functional Approach to a Posteriori Error Estimation for Elliptic Optimal Control Problems with Distributed Control.* Journal of Mathematical Sciences, Vol. 144, No. 6, 2007, pp. 4535–4547

[8] A. Gaevskaya, R.H.W. Hoppe, and S. Repin. *Functional Majorants for Distributed Optimal Control Problems with Control Constraints* Mathematisches Forschungsinstitut Oberwolfach, *Adaptive Numerical Methods for PDEs*, Report 29, 2007, pp. 20–21